

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 14, 2004, 16:55:45 ; Search time 557.336 Seconds
(without alignments)
3959.005 Million cell updates/sec

Title: US-09-807-802A-15

Perfect score: 3251

Sequence: 1 TAPGKRPVRSQPEPSSS.....NNGLYTEPRIGTRYLRPL 599

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_apool/US09807802/runat_11022004_175610_15987/app_query.fasta_1.2389
-DB=Published Applications NA -QPMF=fastap -SUFFX=rnpb -MINMATCH=0.1
-LOOFCLO=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09807802 @CGN_1_11100 @runat_11022004_175610_15987
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-Fgapop=6 -Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
|------------|-------|-------------|--------|-------|-------------|

ALIGNMENTS

RESULT 1

US-10-291-583-6
; Sequence 6, Appl
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alvira, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; FILE REFERENCE: UPN-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; CURRENT FILING DATE: 2002-11-12
; PRIOR FILING DATE: 2002-11-12
; PRIOR FILING DATE: 2001-11-13
; PRIOR FILING DATE: 2001-11-13
; PRIOR FILING DATE: 2001-12-17
; PRIOR FILING DATE: 2001-12-17
; PRIOR FILING DATE: 2002-05-01
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6

| | | | | | | |
|----|--------|-------|------|----|-------------------|--------------------|
| 1 | 3251 | 100.0 | 4718 | 13 | US-10-291-583-6 | Sequence 6, Appl |
| 2 | 3251 | 100.0 | 7447 | 13 | US-10-216-870-11 | Sequence 11, Appl |
| 3 | 2820.5 | 86.6 | 3075 | 13 | US-10-291-583-26 | Sequence 26, Appl |
| 4 | 2815 | 86.6 | 4726 | 13 | US-10-291-583-8 | Sequence 8, Appl |
| 5 | 2808.5 | 86.4 | 3142 | 13 | US-10-291-583-25 | Sequence 25, Appl |
| 6 | 2801 | 86.2 | 8179 | 15 | US-10-205-942-5 | Sequence 5, Appl |
| 7 | 2764.5 | 85.0 | 4679 | 9 | US-09-804-898-1 | Sequence 1, Appl |
| 8 | 2764.5 | 85.0 | 4679 | 9 | US-09-945-681-10 | Sequence 10, Appl |
| 9 | 2764.5 | 85.0 | 4679 | 13 | US-10-136-819-6 | Sequence 6, Appl |
| 10 | 2764.5 | 85.0 | 4679 | 14 | US-10-038-972A-12 | Sequence 12, Appl |
| 11 | 2761.5 | 84.9 | 2205 | 13 | US-10-291-583-120 | Sequence 120, Appl |
| 12 | 2753.5 | 84.7 | 3122 | 13 | US-10-291-583-56 | Sequence 56, Appl |
| 13 | 2751.5 | 84.6 | 3123 | 13 | US-10-291-583-54 | Sequence 54, Appl |
| 14 | 2749.5 | 84.6 | 3113 | 13 | US-10-291-583-55 | Sequence 55, Appl |
| 15 | 2735.5 | 84.1 | 3123 | 13 | US-10-291-583-57 | Sequence 57, Appl |
| 16 | 2735 | 84.1 | 3122 | 13 | US-10-291-583-45 | Sequence 45, Appl |
| 17 | 2726 | 83.9 | 3117 | 13 | US-10-291-583-43 | Sequence 43, Appl |
| 18 | 2722.5 | 83.7 | 4680 | 13 | US-10-375-777-1 | Sequence 1, Appl |
| 19 | 2722.5 | 83.7 | 4680 | 14 | US-10-077-294-1 | Sequence 1, Appl |
| 20 | 2722.5 | 83.7 | 4680 | 14 | US-10-163-886-1 | Sequence 1, Appl |
| 21 | 2722.5 | 83.7 | 4680 | 15 | US-10-263-127-1 | Sequence 44, Appl |
| 22 | 2722 | 83.7 | 3121 | 13 | US-10-291-583-44 | Sequence 44, Appl |
| 23 | 2719 | 83.6 | 3122 | 13 | US-10-291-583-42 | Sequence 42, Appl |
| 24 | 2715 | 83.5 | 3128 | 13 | US-10-291-583-27 | Sequence 27, Appl |
| 25 | 2715 | 83.5 | 3197 | 13 | US-10-291-583-29 | Sequence 29, Appl |
| 26 | 2710 | 83.4 | 3129 | 13 | US-10-291-583-59 | Sequence 59, Appl |
| 27 | 2709.5 | 83.3 | 4721 | 13 | US-10-291-583-1 | Sequence 1, Appl |
| 28 | 2707 | 83.3 | 3123 | 13 | US-10-291-583-41 | Sequence 41, Appl |
| 29 | 2706.5 | 83.3 | 4675 | 10 | US-09-782-378A-1 | Sequence 1, Appl |
| 30 | 2706.5 | 83.3 | 4675 | 10 | US-09-782-378A-2 | Sequence 2, Appl |
| 31 | 2706.5 | 83.3 | 4675 | 13 | US-10-291-583-7 | Sequence 7, Appl |
| 32 | 2706.5 | 83.2 | 4385 | 15 | US-10-240-198-1 | Sequence 5, Appl |
| 33 | 2704.5 | 83.2 | 4385 | 13 | US-10-291-583-5 | Sequence 9, Appl |
| 34 | 2703 | 83.1 | 3098 | 13 | US-10-291-583-9 | Sequence 47, Appl |
| 35 | 2703 | 83.1 | 3128 | 13 | US-10-291-583-47 | Sequence 48, Appl |
| 36 | 2702 | 83.1 | 3128 | 13 | US-10-291-583-28 | Sequence 28, Appl |
| 37 | 2701.5 | 83.1 | 3106 | 13 | US-10-291-583-34 | Sequence 34, Appl |
| 38 | 2700 | 83.0 | 3084 | 13 | US-10-291-583-39 | Sequence 39, Appl |
| 39 | 2699 | 83.0 | 3121 | 13 | US-10-291-583-11 | Sequence 11, Appl |
| 40 | 2698 | 83.0 | 2489 | 13 | US-10-291-583-35 | Sequence 35, Appl |
| 41 | 2696.5 | 82.9 | 3127 | 13 | US-10-291-583-15 | Sequence 15, Appl |
| 42 | 2696.5 | 82.9 | 3276 | 13 | US-10-291-583-38 | Sequence 38, Appl |
| 43 | 2696 | 82.9 | 2495 | 13 | US-10-291-583-36 | Sequence 36, Appl |
| 44 | 2696 | 82.9 | 3098 | 13 | US-10-291-583-37 | Sequence 37, Appl |
| 45 | 2695 | 82.9 | 3106 | 13 | US-10-291-583-18 | Sequence 18, Appl |

```

; LENGTH: 4718
; TYPE: DNA
; ORGANISM: adeno-associated virus serotype 1
US-10-291-583-6

Alignment Scores:
Pred. No.: 0 Length: 4718
Score: 3251.00 Matches: 599
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 13 Gaps: 0

US-09-807-802A-15 (1-599) x US-10-291-583-6 (1-4718)

QY 1 ThrAlaProGlyValLysArgProValGluInSerProGlnInLuproAaspSerSer 20
DB 2634 ACGGCTCCTGGAAAGAAAGCTCCGGTAGACAGTCGCCACAAGACCCAGACTCCTCTCG 2693
QY 21 GlyIleGlyThrGlyGlnGlnProAlaLysLysArgLeuAenPheGlyGlnThrGly 40
DB 2694 GGCATCGGCAAGACAGCCGAGCCCGCTAAAGAGAGACTCAATTTTGGTCAGACTGGC 2753
QY 41 AspSerGluSerValProAaspProGlnProLeuGlyGluProProAlaThrProAlaLa 60
DB 2754 GACTCAGATCAGTCCCGATCCACNACCTCTCGAGAGAACCTTCAGCAACCCCGCTGCT 2813
QY 61 ValGlyProThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAaspAenGluGly 80
DB 2814 GTGGGACCTACTACATGCTTCAGCGGTGGCGCACCAATGCGCAGACAATAACGAAGGC 2873
QY 81 AlaAaspGlyValGlyAenAlaSerGlyAenTrpHisCysAaspSerThrTripleGlyAap 100
DB 2874 GCCGACGGAGTGGTAATGCCCTCAGGAATTTGGCATTTGGCATTCACATGCTGGCGGAC 2933
QY 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTyrAenAenHisLeuTyr 120
DB 2934 AGAGTCATCACCACCGACCGACCTGGGCTTGGCCCACTCAATTAACCACTCTAC 2993
QY 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAenAaspAenHisTyrPheGlyTyrSer 140
DB 2994 AAGCAATCTCCAGTCTTCAACGGGGGCGCAGCAACCACTACTTCCGGCTACAGC 3053
QY 141 ThrProTriGlyTyrPheAaspPheAenArgPheHisCysHisPheSerProArgAapTrp 160
DB 3054 ACCCCCTGGGGGTATTTGATTTCAACAGATTCCTGCTGCTTTCCACCACTGACTGG 3113
QY 161 GlnArgLeuIleAenAenAenTrpGlyPheArgProLysArgLeuAenPheLysLeuPhe 180
DB 3114 CAGCGACTCATCAACAACAAATTTGGGATTCGGGCCCAAGAGACTCAACTTCAAACTCTC 3173
QY 181 AenIleGlnValLysGluValThrThrAenAaspGlyValThrIleAlaAenAenLeu 200
DB 3174 AACATCAAGTCAAGAGAGTCCAGCAATGATGATGGCGTCACCACTCAATCAATCACTT 3233
QY 201 ThrSerThrValGlnValPheSerAapSerGluTyrGlnLeuProTyrValLeuGlySer 220
DB 3234 ACCAGACCGGTTCAAGTCTTCGGACTCGAGTACCACTTCCGTAAGTCTCTCGGCTCT 3293
QY 221 AlaHisGlnGlyCysLeuProProPheProAlaAaspValPheMetIleProGlnTyrGly 240
DB 3294 GCGCACCGGCTGCTCCTCGTTCCCGGGGAGCTGTTATGATTCGCAATCCGCAATACGCG 3353
QY 241 TyrLeuThrLeuAenAenGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
DB 3354 TACCTGACGCTCAACAATGGCAGCCCAAGCGGTGGAGCTTCATCTCTTTTACTGCTGGAA 3413
QY 261 TyrPheProSerGlnMetLeuArgThrGlyAenAenPheThrPheSerTyrThrPheGlu 280
DB 3414 TATTTCCCTTCTCAGATGCTGAGACGGGCAACACTTTTACCTTCAGCTACACCTTTGAG 3473
QY 281 GluValProPheHisSerTyrAlaHisSerGlnSerLeuAaspArgLeuMetAenPro 300

```

RESULT 2

US-10-216-870-11

; Sequence 11, Application US/10216870

; Publication No. US20030148506A1

; GENERAL INFORMATION:

; APPLICANT: KOTIN, ROBERT M

; APPLICANT: URABE, MASASHI

; APPLICANT: DING, CHUAN-TIAN

; TITLE OF INVENTION: PRODUCTION OF ADENO-ASSOCIATED VIRUS IN INSECT CELLS

; FILE REFERENCE: 402133

; CURRENT APPLICATION NUMBER: US/10/216.870

; CURRENT FILING DATE: 2002-08-13
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 11
 ; LENGTH: 7447
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Synthetic
 US-10-216-870-11

Alignment Scores:

Pred. No.: 0 Length: 7447
 Score: 3251.00 Matches: 599
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 13 Gaps: 0

US-09-807-802A-15 (1-599) x US-10-216-870-11 (1-7447)

| | | | |
|----|------|---|------|
| Qy | 1 | ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSer | 20 |
| Db | 5032 | ACGGCTCTCGAAGAAGACGTCGGTAGAGCAGTCGCCAAGAGCCAGACTCTCTCTCG | 5091 |
| Qy | 21 | GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGly | 40 |
| Db | 5092 | GGCATCGGCAAGACAGCGCCAGCAGCCGCTAAAGAGACTCAATTTTGGTCAGACTGGC | 5151 |
| Qy | 41 | AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla | 60 |
| Db | 5152 | GACTCAGAGTCAGTCCCGCGATCCCAACCTCTCGAGAGAACCTCCAGCAACCCCGCTGCT | 5211 |
| Qy | 61 | ValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGly | 80 |
| Db | 5212 | GTGGAGCTACTACAAATGGCTTCAGCGGTGGCGCACCAATGGCAGACAAATACGAGGC | 5271 |
| Qy | 81 | AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp | 100 |
| Db | 5272 | GCCGACGGAGTGGGTAAATGCCCTCAGGAAATGGCAATTCGCATTCACATGCTGGCGAC | 5331 |
| Qy | 101 | ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrThrAsnAsnHisLeuTyr | 120 |
| Db | 5332 | AGAGTCATCACACACAGACCGCGACCTGGGCGCTTGGCCACCTACAAATACCACTCTAC | 5391 |
| Qy | 121 | LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer | 140 |
| Db | 5392 | AAGCAAAATCTCCAGTGGCTTCAACGGGGGCCAGCAACAGCACTACTTCCGCTACAGC | 5451 |
| Qy | 141 | ThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProAspTrp | 160 |
| Db | 5452 | ACCCCTCGGGGTATTTTGAATTTCAACAGATTCACACTGCCACTTTTCCACCGTGACTGG | 5511 |
| Qy | 161 | GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe | 180 |
| Db | 5512 | CAGCGACTCATCAACAAATTTGGGGATTCGGGCCCAAGAGACTCAACTTCAAACTCTTC | 5571 |
| Qy | 181 | AsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeu | 200 |
| Db | 5572 | AACATCCAGTCAAGAGGTCAAGCAAGATGATGGCGTCACNACCATCGCTAATAACCTT | 5631 |
| Qy | 201 | ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer | 220 |
| Db | 5632 | ACCAGCAGGTTCAGTCTCTCTCGACTCGGAGTACCAGTTCGCTCGCTCTCGGCTCT | 5691 |
| Qy | 221 | AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGly | 240 |
| Db | 5692 | GGCGACCAAGGGCTGCTCTCTCTCGCTTCCCGCGGACGTTTCATGATTCCCGCAATACGGC | 5751 |
| Qy | 241 | TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu | 260 |
| Db | 5752 | TACCTGACGCTCAACAAATGGGACGCCAAGCGGTGGGACGTTCTATCTTTTACTGCTGGAA | 5811 |

| | | | |
|----|------|---|------|
| Qy | 261 | TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu | 280 |
| Db | 5812 | TATTTCCCTTCTCAGATGCTGAGAACGGGCAACAACATTTACCTTCAGCTACACCTTTGAG | 5871 |
| Qy | 281 | GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro | 300 |
| Db | 5872 | GAAAGTGCCTTTTCCACAGCAGCTACGGCAGCAGCCGCGACCGCTGATGAATCCT | 5931 |
| Qy | 301 | LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln | 320 |
| Db | 5932 | CTCATCGACCAATACCTGTATTACCTGAAACAGAACTCAAAATCAGTCGGGAAGTGGCAA | 5991 |
| Qy | 321 | AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn | 340 |
| Db | 5992 | AACAAGGACTTGTGTTTAGCGGTGCTCCAGCTGGCATGTGTGTTCAGCCCAAAAC | 6051 |
| Qy | 341 | TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn | 360 |
| Db | 6052 | TGGCTACTGGACCTGTTATCGGCAGCAGCGCTTCTAAACAAACAAACAGACAAAC | 6111 |
| Qy | 361 | AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle | 380 |
| Db | 6112 | AACAGCAATTTTACCTGGACTGGTCTTCAAAATATTAACCTCAATGGCGTGAATCCATC | 6171 |
| Qy | 381 | IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet | 400 |
| Db | 6172 | ATCAACCTCGCACTGCTATGGCTCACACAAAGCAGCAGCAAGAAAGTGTCTTCCCATG | 6231 |
| Qy | 401 | SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAsnAsn | 420 |
| Db | 6232 | AGCGTGTCTATGATTTTGGAAAGAGAGCGCGGAGCTTCAAAACACTGCATTTGGACAT | 6291 |
| Qy | 421 | ValMetIleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe | 440 |
| Db | 6292 | GTCAATGATTACAGCAGGAAGAGGAAATTAAGCCACTAAACCTGTGGCCACCGAAGATTT | 6351 |
| Qy | 441 | GlyThrValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHis | 460 |
| Db | 6352 | GGGACCTGGCAGTCAATTTCCAGAGCAGCAGCAGCAGCCTGCGACCGGAGATGTGCAT | 6411 |
| Qy | 461 | AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro | 480 |
| Db | 6412 | GCTATGGAGCATTTACTTGGCATGGTGTGGCAAGATAGACGCTGACCTGCAGGGTCCC | 6471 |
| Qy | 481 | IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly | 500 |
| Db | 6472 | ATTTGGGCCAAAATTCCTCACAGATGGACACTTTTCCCGCTCTCTCTTATGGCGGC | 6531 |
| Qy | 501 | PheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsn | 520 |
| Db | 6532 | TTTGGACTCAAGAACCCGCTCTCTCAGATCCTCATCAAAAACACGCTGTCTCTCGCAAT | 6591 |
| Qy | 521 | ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly | 540 |
| Db | 6592 | CCTCGCGGAGTTTTCAGCTACAAAGTTTGTCTTATTCATCCACCAATACTTCCACGGA | 6651 |
| Qy | 541 | GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPro | 560 |
| Db | 6652 | CAAGTGAGTGTGGAATTTGAATGGAGCTGCAGAAACAAACAGCAAGCGCTGGAATCCC | 6711 |
| Qy | 561 | GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsn | 580 |
| Db | 6712 | GAAATGTCAGTACATCAATTAATGCAAAATCTGCCAACGTTGATTTTACTGTGGACAAAC | 6771 |
| Qy | 581 | AsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu | 599 |
| Db | 6772 | AATGGACTTTTATCTGAGCTCGCCCTTGGCACCCGCTTACCTTACCTGCTCCCTG | 6828 |

RESULT 3

US-10-291-583-26
 ; Sequence 26, Application US/10291583
 ; Publication No. US2003013877A1
 ; GENERAL INFORMATION:

; APPLICANT: Gao, Guangping
 ; APPLICANT: Wilson, James M.
 ; APPLICANT: Alvira, Mauricio
 ; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus (AAV) Sequences
 ; FILE REFERENCE: UPN-02733USA
 ; CURRENT APPLICATION NUMBER: US/10/291,583
 ; CURRENT FILING DATE: 2002-11-12
 ; PRIOR APPLICATION NUMBER: US 60/350,607
 ; PRIOR FILING DATE: 2001-11-13
 ; PRIOR APPLICATION NUMBER: US 60/341,117
 ; PRIOR FILING DATE: 2001-12-17
 ; PRIOR APPLICATION NUMBER: US 60/377,066
 ; PRIOR FILING DATE: 2002-05-01
 ; PRIOR APPLICATION NUMBER: US 60/386,675
 ; PRIOR FILING DATE: 2002-06-05
 ; NUMBER OF SEQ ID NOS: 120
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 26
 ; LENGTH: 3075
 ; TYPE: DNA
 ; ORGANISM: new AAV serotype, clone H2
 US-10-291-583-26

Alignment Scores:
 Pred. No.: 8,288-298 Length: 3075
 Score: 2820.50 Matches: 506
 Percent Similarity: 91.32% Conservative: 41
 Best Local Similarity: 84.47% Mismatches: 51
 Query Match: 86.76% Indels: 1
 DB: 13 Gaps: 1

US-09-807-802A-15 (1-599) x US-10-291-583-26 (1-3075)

QY 1 ThrAlaProGlyLysLysArgProValGluInSerProGluInProAspSerSerSer 20
 DB 1214 ACGGCTCCGGGAAAAAGAGGCGGTAGACACTCTCTGTGGAGCCAGACTCTCTCTCG 1273
 QY 21 GlyIleGlyThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGly 40
 DB 1274 GGAACCGGAAACGCGGCGGCGCTGCAGAAAAAGATTAAATTTTGGTCAGACTGGA 1333
 QY 41 AspSerGluSerValProAspProGluInProLeuGlyGluProProAlaThrProAla 60
 DB 1334 GAGCAGACTCTCGTACTGACCCCGAGCTCTCGACAGCCACAGCAGCCCTCTGGT 1393
 QY 61 ValGlyProThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAspAsnGluGly 80
 DB 1394 CTGGGATCTACTACATGCTACAGCGAGTGGCGGACCAATGGCAGACAAATAACGAGGT 1453
 QY 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
 DB 1454 GCGGATGGAGTGGGTAAATTCCTCAGAAAAATGGCATTTCCCAATGGCTGGGCGAC 1513
 QY 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrThrAsnAsnHisLeuTyr 120
 DB 1514 AGAGTCATCACACACAGCAGCCGAACTGGGCGCTGCCACATACAAACACCCCTCTAC 1573
 QY 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAsnHisTyrPheGlyTyrSer 140
 DB 1574 AAGCAATCTCCAGCCAAATCA---GGAGCGCAGCAACGACCACTACTCTGGCTACAGC 1630
 QY 141 ThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
 DB 1631 ACCCTCGGGGTATTTTGATTTCAACAGATTCCACTGGCCCTTTTCCACAGTGACTGG 1690
 QY 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
 DB 1691 CAAGACTCATCAACCAACACTGGGGATTCGGSCCCCAAGACTCAACTTCAAGCTCTTT 1750
 QY 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnLeu 200
 DB 1751 AATATTCAAGTCAAGAGGTTCAGCGAATGACGAGTACGCGTACGAGCAAGTTCGCAATACCTTT 1810

QY 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
 DB 1811 ACCAGCAGCGGTTCAGGTGTTTACTGACTCGAGTACGAGTCCGCTACGCTCGGGCTCG 1870
 QY 221 AlaHisGlnGlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGly 240
 DB 1871 GCGCATCAAGGATGCTCCCGCGTTTCAGCGGAGCTCTTCATGCTCCACAGTAGTGA 1930
 QY 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
 DB 1931 TACCTCACCTGAAACCAACGAGTTCAGGCGGTAGGACGCTCTTCTCTTTTATCTGCTGAG 1990
 QY 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
 DB 1991 TACTTCTCTCTCAGATGCTGCGTACTGGAACACACTTTCAGTTCAGTACACTTTTGA 2050
 QY 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
 DB 2051 GAGTGCCTTTCCACAGCAGCTACGCTCACAGCAGAGTCTGGATCGGCTGATGAATCCT 2110
 QY 301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
 DB 2111 CTGATCAGCAGTACTCTGATTATCTGAACAGACACAAATAGTGAACCTCTTCAG 2170
 QY 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
 DB 2171 CAGTCTCGGTACTGTTTACCAAGCTGGACCAACCACTGCTCTTCAAGCTAATAAC 2230
 QY 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
 DB 2231 TGGCTGCTGAGCTTCTGTACAGCAGCAGCGTCTGTCAAACAGGCAACAGCAACAAAC 2290
 QY 361 AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 380
 DB 2291 AACGCAACTTCTCTCGGACTGACAGTACAAAGTATCATCTAAATGGCGGAGCTCGTGT 2350
 QY 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet 400
 DB 2351 GTTAAATCCAGGACGAGCTATGGCCAGTCACAGGATGACGAGAAAGAAAGTTTTCCTCC 2410
 QY 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
 DB 2411 CATGGAACCTGTATTTGGTAAACAGGAAACAAATGCCAACGACGCGGATTTTGGAAAT 2470
 QY 421 ValMetIleThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
 DB 2471 GTCATGATTACAGATGAAGAAAGAAATCAGGGCCCACTCCCTGGCTTACGAGCAGTAC 2530
 QY 441 GlyThrValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHis 460
 DB 2531 GGGACTGTGTCAATAATTTGCAAACTCAAACTGGTCCCACTACTTGGAACTGTCAAT 2590
 QY 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro 480
 DB 2591 CGCCAAGAGGCTTACCTGTGTGTGGCAGGATCGAGACGTGTACTCGAGGAGCCC 2650
 QY 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
 DB 2651 ATTTGGGCAAGATTCCTCACCGATGACACTTTTCTCTCTCTTCTCTCTCTGAGGT 2710
 QY 501 PheGlyLeuLysAsnProProGlnIleLeuLysAsnThrProValProAlaAsn 520
 DB 2711 TTTGAGCTCAAAACCCGCTCTCAGATCATGATCAAAACCACTCCCGCTTCCAGCAAT 2770
 QY 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlyTyrSerThrGly 540
 DB 2771 CTCTCCCAAACTTTCAGTTCTGCCAAGTTTCTCTTCTCATCACACAGTATTTCCAGGGA 2830
 QY 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPro 560
 DB 2831 CAGGTACGCTGGAGATCGATGGGAGCTGCAGAGGAGGACAGCAACAGCTGGATCCC 2890

Qy 561 GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsn 580
Db 2991 GAAATTCAGTACACTTCCAACTACACAGTCTGTATATGGGACTTTACTGTGGACACT 2950
Qy 581 AsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db 2951 AATGGTGTATTACAGAGCTCGCCCATTCGGCCAGATACCTGACTCGTAATCTG 3007

RESULT 4
US-10-291-583-8
; Sequence 8, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wileon, James M.
; APPLICANT: Alvira, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; FILE REFERENCE: UPN-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 8
; LENGTH: 4726
; TYPE: DNA
; ORGANISM: adeno-associated virus serotype 3
US-10-291-583-8

Alignment Scores:
Pred. No.: 6,476-297 Length: 4726
Score: 2815.00 Matches: 511
Percent Similarity: 91.00% Conservative: 35
Best Local Similarity: 85.17% Mismatches: 52
Query Match: 86.59% Indels: 2
DB: 13 Gaps: 2

US-09-807-802A-15 (1-599) x US-10-291-583-8 (1-4726)

Qy 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSer 20
Db 2620 ACGGCTCTGGAAAGAGGGGCTGTAGATCAGTCTCTCAGGAACCGGACTCATCT 2679
Qy 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGly 40
Db 2680 GGTGTGGCAATCGGGCAACAGCTCGCCAGAAAGAACTAAATTTTCGGTCAGACTGGA 2739
Qy 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
Db 2740 GACTCAGAGTCTGCTCCAGACCTCACTCTCGAGAACCCAGGAGGCCCCCACAAGT 2799
Qy 61 ValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGly 80
Db 2800 TTGGGATCTAATACAAATGGCTTCAGGCGGTGGCGCCCAATGGCAGACAATAACGAGGT 2859
Qy 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTphHisCysAspSerThrTyrLeuGlyAsp 100
Db 2860 GCCGATGAGTGGGTAATTCCTCAGGAATTCGCAATTCGATTCCTCAATGGCTGGCGGAC 2919
Qy 101 ArgValIleThrThrSerThrArgThrTyrAlaLeuProThrTyrAsnAsnHisLeuTyr 120
Db 2920 AGAGTCATCACCACGACGACGACGACCTGGGCGCTGCCACCTTACAAACCACTCTTAC 2979
Qy 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer 140

Db 2980 AAGCAAAATCTCCAGCCCAATCA---GGAGCTTCAAACGACCAACCACTACTTTGGCTACAGC 3036
Qy 141 ThrProTyrGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTyr 160
Db 3037 ACCCTTGGGGGTATTTTGACTTTTACAGATTTCATTCGCCACTTCTCACCACTGACTGG 3096
Qy 161 GlnArgLeuIleAsnAsnAsnTyrGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
Db 3097 CAGCGACTCATTAACAACAACCTGGGATTCGGGCCCAAGAAACCTCAGCTTCAAGCTCTTC 3156
Qy 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeu 200
Db 3157 AACATCCAGTTAGAGGGGTTCACGCAAGACGATGGGACGACGACTATTTGCCAATAACCTT 3216
Qy 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
Db 3217 ACAGACAGGTTCAAGTGTTCAGGACTCGAGATTCAGCTCCCGTACGTCGCGGTGCG 3276
Qy 221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGly 240
Db 3277 GCGCACCAGGCTGTCTCCGGCGTTTCCAGCGGACGCTTTCATGGTCCCTCAGTAGTGA 3336
Qy 241 TyrIleuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerPheTyrCysLeuGlu 260
Db 3337 TACCTCACCTGAACAACGGAAGTCAAGCGGTGGGACGCTCATCTTTTACTGCTCGAG 3396
Qy 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
Db 3397 TACTTCCCTTCGAGATGCTAAGGACTGGAAATTAATCTTCCAATTCAGTATACCTTCGAG 3456
Qy 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
Db 3457 GATGTACTCTTTTCCAGCAGCTACGCTCACAGCGAGTGGATTTGGATCGCTTGAATCCT 3516
Qy 301 LeuIleAspGlnTyrIleuTyrTyrLeuAsnArgThrGln---AsnGlnSerGlySerAla 319
Db 3517 CTTATTGATCAGTATCTGTACTCTACCTGAAACAGAACGCAACCACTCTGGAAACACC 3576
Qy 320 GlnAsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLys 339
Db 3577 AACCAATCAGCGTCTCTTTTAGCCAGGCTGGGCTCAGTCTATGCTTTTGCAGGCCAGA 3636
Qy 340 AsnTyrLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsn 359
Db 3637 AATTGGCTACTCGGCGCTCTACCGCAACAGAGACTTTCAAAGACTGCTTAACGACAAC 3696
Qy 360 AsnAsnSerAsnPheThrTyrThrGlyAlaSerLysTyrAsnLeuAsnGlyValGluSer 379
Db 3697 AACCAACAGTAACCTTCTTGGACAGCGCGCCAGCAAAATATCATCTCAATGGCCGCGACTCG 3756
Qy 380 IleIleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePhePro 399
Db 3757 CTGGTGAATCCAGGACGAGCTATGGCCAGTACACAGGACGATGAAGAAATAATTTTCCCT 3816
Qy 400 MetSerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAsp 419
Db 3817 ATGCAGCGCAATCTAATATTGGCAAGAGGCAACACGCGCAGTAACGCAAGATTAAGAT 3876
Qy 420 AsnValMetIleThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArg 439
Db 3877 AATGTAATGATTCAGGATGAAGAGATTCGTATACCACCAATCTCTGTGGCAACAGGCG 3936
Qy 440 PheGlyThrValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspVal 459
Db 3937 TATGGAACTGTGGCAATTAATTCGAGAGCTCAATACAGCTCCACGACTGGAAATGTC 3996
Qy 460 HisAlaMetGlyAlaLeuProGlyMetValTyrGlnAspArgAspValTyrLeuGlnGly 479
Db 3997 AATCATCAGGGGCGCTTACCTGGCATGCTGTGGCAAGATCGTGACGTGTACTCTTCAAGGA 4056
Qy 480 ProIleTyrAlaValIleProHisThrAspGlyHisPheHisProSerProLeuMetGly 499
Db 4057 CCTATCTGGGCAAGATTCCTCTCACACGAGTGGACACTTTTCACTCTCTCTGTATGGGA 4116

```
QY 500 GlyPheGlyLeuLysAsnProProGlnIleLeuLysAsnThrProValProAla 519
Db 4117 GCGTTTGAGTGAACATCCGCTCTCAATCATGATCAAAATACATCCGGTACCGCA 4176
QY 520 AnProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThr 539
Db 4177 AATCTCTCGACGACTTTCAGCCCGCCAGTTTGTCTCATTTATCAGTACTCCACT 4236
QY 540 GlyGlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsn 559
Db 4237 GCACAGGTGAGTGGGATTCAGTGGGAGCTACAGAAAGAAACAGCAAAAGTTGGAT 4296
QY 560 ProGluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAsp 579
Db 4297 CCAGAGATTTCAGTACTTCCAACTACAAAGTCTGTTAATGTGGACTTTACTGTAGAC 4356
QY 580 AnAsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db 4357 ACTAATGGTGTATTAGTGAACCTCGCCCTATTGGAAACCCGTAATCTACACGAAACTTG 4416

RESULT 5
US-10-291-583-25/c
; Sequence 25, Application US/10291583
; Publication No. US2003013872A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alvira, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; TITLE OF INVENTION: Sequences and Isolating No. US2003013872A1e1 Sequences Identifi
; FILE REFERENCE: UPN-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; PRIOR FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; TYPE: DNA
; ORGANISM: new AAV serotype, clone H6
US-10-291-583-25

Alignment Scores:
Pred. No.: 1,76e-296 Length: 3142
Score: 2808.50 Matches: 505
Percent Similarity: 91.15% Conservative: 41
Best Local Similarity: 84.31% Mismatches: 52
Query Match: 86.39% Indels: 1
DB: 13 Gaps: 1

US-09-807-802A-15 (1-599) x US-10-291-583-25 (1-3142)
QY 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSer 20
Db 1928 ACGCTCCGGGAAAAAGAGCGGTAGACACTCTCTGTGGAGCCAGACTCTCTCTCG 1869
QY 21 GlyIleGlyThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGly 40
Db 1868 GGAACCGGAAAGCGCGCAGCAGCTGCAAGAAAGAAATTAATTTTGGTCTGAGCTGA 1809
QY 41 AppSerGluSerValProAspProGlnProLeuGlyGluProProAlaAla 60
Db 1808 GAGCGAGATCCGTACTACCCCGCCCTCTCGACACGCCACCGACGCCCTCTGTT 1749
QY 61 ValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnGluGly 80
Db 1748 CTGGGATCTACTACATGGCTACAGCGAGTGGCCCAATATGGCAGCAATATACAGGGGT 1689
QY 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
Db 1688 GCGGATGGAGTGGGTAAATCTCTCAGGAAATGGCATTTGCCATTTCCCAATGGCTGGCGAC 1629
QY 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyr 120
Db 1628 AGAGTCTATCACACACGACCCGAACTGGGCCCTCGCCCATACATACAAACACCTCTAC 1569
QY 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer 140
Db 1568 AAGCAATATCTCCAGCCCATCA--GGAGCCAGCAACGACCACTACTTTGGCTACAGC 1512
QY 141 ThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
Db 1511 ACCCCCTGGGGTATTTTGAATTCACAGATTCCTCACTGCCACTTTTCCACACGCTGACTGG 1452
QY 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
Db 1451 CAAAGACTCGTCAACAAACACCGGGGATTCGGGCCCAAAAGACTCAACTTCAAGCTCTTT 1392
QY 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeu 200
Db 1391 AATATTCAAGTCAAGAGGTTCACGAGATGAGCGTACGACGAGATTTGCCAATAACCTT 1332
QY 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
Db 1331 ACCAGCAGGTTCAGGTGTTTACTGACTCGGAGTACGAGCTCCGCTCGCTCGGCTCG 1272
QY 221 AlaHisGlnGlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGly 240
Db 1271 GCGCATCAAGGATGCTCCCGCCGTTTCCAGCGGAGCTCTTCATGGTCCCAAGATATGGA 1212
QY 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
Db 1211 TACCTCACCTGTAACACGCGGAGTCAGCGGTAGGAGCTCTCTCTTTTACTGCTGGAG 1152
QY 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
Db 1151 TACTTTCTCTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1092
QY 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
Db 1091 GACGTGCTTTCACAGCAGCTACCTCAGCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCT 1032
QY 301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
Db 1031 CTGATCGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 972
QY 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
Db 971 CAGTCTCGGCTACTGTTTAGCCAACTGAGCAACCAACCACTGCTCTCTCAAGCTAAAC 912
QY 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
Db 911 TGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 852
QY 361 AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 380
Db 851 AACGACAACTTTCCCTGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 792
QY 381 IleAsnProGlyThrAlaMetAlaSerHisLysLysAspGluAspLysPhePheProMet 400
Db 791 GTTAATCCAGACCCAGCTATGCGCAGTACCAAGGATGACGAGAAAGTATTTTCCCATG 732
QY 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
Db 731 CATGGACCCCTGATATTTGGTAACAGGAAACAAATGCGACGCGGATTTGGAAAT 672
QY 421 ValMetIleThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
Db 672 CATGGACCCCTGATATTTGGTAACAGGAAACAAATGCGACGCGGATTTGGAAAT 612
QY 612 CATGGACCCCTGATATTTGGTAACAGGAAACAAATGCGACGCGGATTTGGAAAT 552
Db 552 CATGGACCCCTGATATTTGGTAACAGGAAACAAATGCGACGCGGATTTGGAAAT 492
QY 492 CATGGACCCCTGATATTTGGTAACAGGAAACAAATGCGACGCGGATTTGGAAAT 432
Db 432 CATGGACCCCTGATATTTGGTAACAGGAAACAAATGCGACGCGGATTTGGAAAT 372
QY 372 CATGGACCCCTGATATTTGGTAACAGGAAACAAATGCGACGCGGATTTGGAAAT 312
Db 312 CATGGACCCCTGATATTTGGTAACAGGAAACAAATGCGACGCGGATTTGGAAAT 252
QY 252 CATGGACCCCTGATATTTGGTAACAGGAAACAAATGCGACGCGGATTTGGAAAT 192
Db 192 CATGGACCCCTGATATTTGGTAACAGGAAACAAATGCGACGCGGATTTGGAAAT 132
QY 132 CATGGACCCCTGATATTTGGTAACAGGAAACAAATGCGACGCGGATTTGGAAAT 72
Db 72 CATGGACCCCTGATATTTGGTAACAGGAAACAAATGCGACGCGGATTTGGAAAT 12
```

Db 671 GTCATGATTACAGATGAAGAATAATCAGGCCACCATAATCCCGTGGCTACCGAGCAGTAC 612
Qy GlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHis 460
Db 611 GGGACTGTGTCAATAATTTGCNAAACTCAAACTGGTCCCACTACTTGGAACTGTCAAT 552
Qy AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTrpLeuGlnGlyPro 480
Db 551 CACCAAGAGCGGTACCTGGTATGGTGGCAGGATCGAGACGTGTACTCTGACGGGACCC 492
Qy IleTrpAlaValIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
Db 491 ATTGGGCCAAGATTCCTCACACCAGTGGACACTTTTCATCTTCCCACTGATGGAGGT 432
Qy PheGlyLeuLysAsnProProGlnIleLeuLysAsnThrProValProAlaAsn 520
Db 431 TTTGGACTCAAAACCCGCTCTCCAGATCATGATCAAAAACACTCCCGTTCAGCCCAAT 372
Qy ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTrpSerThrGly 540
Db 371 CCTCCCAAAACTTCAGTTCTGCCAAGTTTGTCTTTCATCACACAGTATTCACGGGA 312
Qy GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPro 560
Db 311 CAGGTCAGCGTGAGATCGAGTGGGAGCTGCAGAGGAGAACAGCAACGCTGGAAATCCC 252
Qy GluValGlnTrpThrSerAsnTrpAlaLysSerAlaAsnValAspPheThrValAspAsn 580
Db 251 GAAATTCAGTACACTTCCAACTACAAACAGTCTGTTAATGTGGACTTTTACTGTGGACACT 192
Qy AsnGlyLeuTrpThrGluProArgProIleGlyThrArgTrpLeuThrArgProLeu 599
Db 191 AATGGTGTATTACAGCGCTCGCCCAATGGCCACAGATACCTGACTCGTAATCTG 135

RESULT 6

US-10-205-942-5
; Sequence 5, Application US/10205942
; Publication No. US20030053990A1
; GENERAL INFORMATION:
; APPLICANT: University of No. US20030053990A1th Carolina-Chapel Hill
; APPLICANT: Rabinowitz, Joseph E.
; APPLICANT: Samulek, Richard J.
; APPLICANT: Xiao, Weidong
; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING THE SAME
; FILE REFERENCE: 5470-186
; CURRENT APPLICATION NUMBER: US/10/205,942
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 8179
; TYPE: DNA
; ORGANISM: Adeno-associated virus
US-10-205-942-5

Alignment Scores:
Pred. No.: 5,19e-295 Length: 8179
Score: 2801.00 Matches: 509
Percent Similarity: 90.33% Conservative: 33
Best Local Similarity: 84.83% Mismatches: 55
Query Match: 86.16% Indels: 2
DB: 15 Gaps: 2

US-09-807-802A-15 (1-599) x US-10-205-942-5 (1-8179)

Qy 1 ThrAlaProGlyLysValArgProValGluGlnSerProGlnGluProAspSerSer 20
Db 2544 ACGGCTCCGGGAAAAAAGAGCGCGGTAGACACTCTCTGTGGAGCCAGACTCTCTCCG 2603
Qy 21 GlyIleGlyThrGlyGlnGlnProAlaLysArgLeuAsnPheGlyGlnThrGly 40
Db 2604 GGAACCGGAAGCGGGCCAGCAGCTGCAAGAAAAGATTGAATTTGGTCAGACTGGA 2663

Qy 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
Db 2664 GAGCGCAGCTCAGTACCTGACCCCGCAGCTCTCGGACAGCCACCGAGCGCCCTCTGGT 2723
Qy 61 ValGlyProThrThrMetAlaSerGlyGlyValAlaProMetAlaAspAsnAsnGlyGly 80
Db 2724 CTGGGAACCTAATACGATGGCTACAGGAGTGGCGCCCAATGGCAGACATACAGAGGC 2783
Qy 81 AlaAspGlyValGlyAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
Db 2784 GCCGACGAGTGGGTAAATCTCCGGAAATGGCATTTGCCATTCCACATGATGGCGAC 2843
Qy 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTrpAsnAsnHisLeuTrp 120
Db 2844 AGAGTCATATCACCAACAGCCGACCTGGGCCCTGCCACCTACACAAACACCTCTTAC 2903
Qy 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTrpPheGlyTrpSer 140
Db 2904 AAACAAATTTCCAGCCCAATCA--GGAGCTTCGACACGACATCACTACTTTGGCTACAGC 2960
Qy 141 ThrProTrpGlyTrpPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
Db 2961 ACCCTTTGGGGTATTGTTGACTTCAACAGATTCCACTGCCACTTTTCCACACGTGACTGG 3020
Qy 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
Db 3021 CAAAGACTCATCAACAACTGGGGATTCCGACCCCAAGAGACTCAACTTCAAGCTCTTT 3080
Qy 181 AsnIleGlnValLysGluValThrAsnAspGlyValThrIleAlaAsnAsnLeu 200
Db 3081 AACATTCAGTCAAGAGGTCAACGACATGACGGTACGACGACGATTTGCCAATAACCTT 3140
Qy 201 ThrSerThrValGlnValPheSerAspSerGluTrpGlnLeuProTrpValLeuGlySer 220
Db 3141 ACCAGCAGGTTTCAAGTGTGTTACTGACTCGGAGTACCAGCTCCGCTCGCTCGGGTCG 3200
Qy 221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTrpGly 240
Db 3201 GCGCACCAAGCTGTCTCCCGCCCTTTCCAGCGAGCGCTTTCATGGTCCCTCAGTATGGA 3260
Qy 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTrpCysLeuGlu 260
Db 3261 TACCTCACCTGACAAACGGAAGTCAACGGTGGGACGCTCATCTTTTACTGCTGGAG 3320
Qy 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTrpThrPheGlu 280
Db 3321 TACTTCCCTTCGACAGATGCTAAGGACTGGAAATAACTTCCAATTCAGCTATACCTTCGAG 3380
Qy 281 GluValProPheHisSerSerTrpAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
Db 3381 GATGTACTCTTTTACAGAGCTACGCTCACAGCAGAGTTTGGATCGCTTGTATGAATCCT 3440
Qy 301 LeuIleAspGlnTrpLeuTrpTrpLeuAsnArgThrGln---AsnGlnSerGlySerAla 319
Db 3441 CTTATTGATCAGTATCTGTACTACTGACAGACGCGCAAGCAACACCTCTGGAACACACC 3500
Qy 320 GlnAsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLys 339
Db 3501 AACCAATCACGGCTGCTTTTATAGCAGCGCTGGGCTCAGTCTATGCTCTTTTCGAGCCAGA 3560
Qy 340 AsnTrpLeuProGlyProCysTrpArgGlnGlnArgValSerLysThrAspAsn 359
Db 3561 AATTGGCTTACCTGGGCGCTGCTACCGCAACAGAGACTTTTCAAGAGCTGCTAACGCAAC 3620
Qy 360 AsnAsnSerAsnPheThrTrpThrGlyAlaSerLysTrpAsnLeuAsnGlyArgGluSer 379
Db 3621 AACCAACGATTAATCTTCTTGGACAGCGCCAGCAATAATCATCTCAATGGCGGACTCG 3680
Qy 380 IleIleAsnProGlyThrAlaMetAlaSerHisLysAspGluAspLysPhePhePro 399
Db 3681 CTGGTGAATCCAGGACGACTATGCCAGCTCACAAGGACGATGAAGAAAAATTTTCCCT 3740
Qy 400 MetSerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAsp 419

Db 3741 ATGCAGCGCAATCTAATATTTGGCAAGAGGACACCGCAAGTACCCAGCAATAGAT 3800
Qy 420 AsnValMetIleThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArg 439
Db 3801 AATGTAATGATTACGATCAAGAGAGATTCGTACCACCAATCTGTGGCAACAGAGCAG 3860
Qy 440 PheGlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspVal 459
Db 3861 TATGGAACTGTGGCAAAATTAATTCAGAGCTCAATACAGCTCCACAGACTGGAACTGTC 3920
Qy 460 HisAlaMetGlyAlaLeuProGlyMetValTyrGlnAspArgAspValTyrLeuGlnGly 479
Db 3921 AATCATCAGGGGCTTACTTGGCATGGTGGCAGATCGTGACGTGTACCTTCAGGA 3980
Qy 480 ProIleTTPAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGly 499
Db 3981 CCTATCTGGGCAAGATTCTCACACGGATGGACACTTTCATCTCTCTCTGATGGGA 4040
Qy 500 GlyPheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAla 519
Db 4041 GGCCTTGGAGTGAACATCGCTCTCAATCATGATCAAAATACTCCGGTACCTGCG 4100
Qy 520 AsnProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThr 539
Db 4101 AATCCTTCGACCACCTTCAGTGGCGCAAGTTTGCTTCTTCATCACACAGTACTCCAG 4160
Qy 540 GlyGlnValSerValGluIleGluTyrGluLeuGlnLysGluAsnSerLysArgTyrAsn 559
Db 4161 GGACAGGTCAAGCTGAGATCGAGTGGAGCTGCAGAGGAAACAGCAAAACGCTGGAAT 4220
Qy 560 ProGluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAsp 579
Db 4221 CCGGAATTCAGTACACTTCCAACTACAAAGTCTGTTAATCGTGGACTTACCGTGGAT 4280
Qy 580 AsnAsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db 4281 ACTAATGGCGTGTATTACAGAGCTCGCCCATGGCCACAGATACCTGCTGTAATCTG 4340

RESULT 7

US-09-804-898-1
; Sequence 1, Application US/09804898
; Patent No. US20020045264A1
; GENERAL INFORMATION:
; APPLICANT: DURING, MATTHEW
; APPLICANT: XIAO, WEIDONG
; TITLE OF INVENTION: PRODUCTION OF CHIMERIC CAPSID VECTORS
; FILE REFERENCE: 102182-14
; CURRENT APPLICATION NUMBER: US/09/804,898
; CURRENT FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 60/189,110
; PRIOR FILING DATE: 2000-03-14
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 4679
; TYPE: DNA
; ORGANISM: adeno-associated virus 2
US-09-804-898-1

Alignment Scores:

| | | | |
|------------------------|-----------|---------------|------|
| Pred. No.: | 2,15e-291 | Length: | 4679 |
| Score: | 2764.50 | Matches: | 496 |
| Percent Similarity: | 89.98% | Conservative: | 43 |
| Best Local Similarity: | 82.80% | Mismatches: | 59 |
| Query Match: | 85.04% | Indels: | 1 |
| DB: | 9 | Gaps: | 1 |

US-09-807-802A-15 (1-599) x US-09-804-898-1 (1-4679)

Qy 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSer 20
Db 2614 ACGGCTCCGGGAAAAAGAGCCGGTAGAGCACTCTCCTGTGGAGCAGACTCCTCTCTCG 2673

Qy 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGly 40
Db 2674 GGAACCGGAAAGGGCGGCGCAGCAGCTTCGAAAAAAAGATTGAATTTTGGTCAGACTGGA 2733
Qy 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
Db 2734 GACGACAGACTCAGTACCTGACCCCGAGCTCTCGACAGCCACCCAGCAGCCCTCTGCT 2793
Qy 61 ValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGlyGly 80
Db 2794 CTGGGAATAATACGATCGCTACAGGCACTGGCGCACCACCAATGGCAGACAATAACGAGGC 2853
Qy 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTyrHisCysAspSerThrTyrLeuGlyAsp 100
Db 2854 GCGGACGAGGTGGGTAAATCTTCGGGAATTTGGCATTGGCATTCACATGGATGGCGCAG 2913
Qy 101 ArgValIleThrThrSerThrArgThrTyrAlaLeuProThrTyrAsnAsnHisLeuTyr 120
Db 2914 AGAGTCATCACCCAGCAGCAGCCGAACTGGGCGCTGCCACCCTACACACACCACTCTAC 2973
Qy 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer 140
Db 2974 AAACAAATTTCCAGCAATCA---GGAGCCTCGAAACGACAATCACTACTCTTGGCTACAGC 3030
Qy 141 ThrProTyrGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTyr 160
Db 3031 ACCCTTGGGGTATTTTGACTTCAACAGATTCCTACTGCCACTTTTACCAGCTGACTGG 3090
Qy 161 GlnArgLeuIleAsnAsnAsnTyrGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
Db 3091 CAAGACTCATCAACAACACTGGGGATTCCGACCCCAAGAGACTCAACTTCAAGCTCTTT 3150
Qy 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrIleAsnAsnLeu 200
Db 3151 AACATTCAAGTCAAAAGAGTCCAGCAGAAATGACGGTAGCAGACGATTCGCCAATTAACCTT 3210
Qy 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
Db 3211 ACCAGCAGCGTTACGGTGTACTGCTCGGAGTACAGCTCCCGTCCGTCTCGGCTCG 3270
Qy 221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGly 240
Db 3271 GCGCATCAAGGATGCTCCGCGCTTCCAGCAGACGCTTTCATGTCGACAGATATGGA 3330
Qy 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
Db 3331 TACCTCACCTCGAACAACGGGAGTCAGGAGTAGGACGCTCTTCATTTTACTGCCTGGAG 3390
Qy 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
Db 3391 TACTTTCTCTCAGATGCTGCTACCGGAAACAACATTTTACCTTCAGCTACACTTTTGG 3450
Qy 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
Db 3451 GACGTTCTCTTCCACAGCAGCTACGCTCACAGCCAGAGTCTCGACCGCTCTCATGAATCT 3510
Qy 301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
Db 3511 CTCATCGACCACTGCTGATTACTTTGAGCAGACAACAACACTCCCAAGTGGAGACCACCG 3570
Qy 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
Db 3571 CAGTCAAGGCTTCAGTTTCTCAGCGCGGAGGAGTGACATTCGGGACCAGCTAGGAAC 3630
Qy 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
Db 3631 TGGCTTCTCTGGACCTGTTCACCGCCAGCAGCGAGTATCAAGACATCTCGCGGATACAAC 3690
Qy 361 AsnSerAsnPheThrThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 380
Db 3691 AACAGTGAATACTCGTGGAGCTGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTG 3750

Qy 381 IleAenProGlyThrAlaMetAlaSerHisLysAspGluuAspLysPhePheProMet 400
Db 3751 GTGAATCGGCGCCGCGCATGCAAGCCACCAAGACGATGAAGAAAGTTTTTCCTCAG 3810
Qy 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAenThrAlaLeuAspAsn 420
Db 3811 AGCGGGGTTCATCTTTGGGAGCAAGGCTCAGAGAAACAAATGTGGACATTGGAAG 3870
Qy 421 ValMetIleThrAspGluGluGluLysAlaThrAenProValAlaThrGluuArgPhe 440
Db 3871 GTCATATTACAGACGAGAGAAATCAGCAACCAATCCCGTGGCTACGGACGATAT 3930
Qy 441 GlyThrValAlaValAenPheGlnSerSerThrAspProAlaThrGlyAspValHis 460
Db 3931 GGTTCGTATCTACCAACCTCCAGAGAGCAACAGACGAGCTACCGCAGATGTCAAC 3990
Qy 461 AlaMetGlyAlaLeuProGlyMetValTropGlnAspArgAspValTyrLeuGlnGlyPro 480
Db 3991 ACACAAGCGGTCTTCAGGCGATGCTCGGCAGACAGAGATGTGTACCTTCAGGGGCC 4050
Qy 481 IleTropAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
Db 4051 ATCTGGGCARAGATTCACACAGCGAGCGACATTTTCACCCCTCTCCCTCATCGGTGA 4110
Qy 501 PheGlyLeuLysAenProProGlnIleLeuLysAenThrProValProAlaAen 520
Db 4111 TTCGACTTAACACCCCTCTCCACAGATTTCTCATCAAGAACACCCCGGTACCTGCGAAT 4170
Qy 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
Db 4171 CCTTCGACCCTTCAGTGGCGGAAAGTTTGCTTCTTCATCACACAGATCTCCACGGGA 4230
Qy 541 GlnValSerValGluIleGluTropGluLeuGlnLysGluAenSerLysArgTropAsnPro 560
Db 4231 CAGGTCAAGTGGAGATCGAGTGGAGCTCGAGAAGGAAACAGCAACGCTGGAATCCC 4290
Qy 561 GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAenValAenPheThrValAspAsn 580
Db 4291 GAAATTCAGTACACTTCCAACTACAAACAAAGTCTGTAAATGGGACTTTTACTGTGGAC 4350
Qy 581 AsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db 4351 AATGGCGGTATTCAGACCTCGCCCATTTGGCACCAGATACCTGACTCTGTAATCTG 4407

RESULT 8

US-09-945-681-10
; Sequence 10, Application US/09945681
; Patent No. US2002064878A1
; GENERAL INFORMATION:

; APPLICANT: UNIVERSITE DE NANTES
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR RECOMBINANT
; TITLE OF INVENTION: ADENO-ASSOCIATED VIRUS PRODUCTION
; FILE REFERENCE: B4182AA - UNIVERSITE DE NANTES
; CURRENT APPLICATION NUMBER: US/09/945,681
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: PC/EP 00/01854
; PRIOR FILING DATE: 2000-03-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10

; LENGTH: 4679
; TYPE: DNA
; ORGANISM: adeno-associated virus 2

US-09-945-681-10

Alignment Scores:

| | | | |
|------------------------|-----------|---------------|------|
| Pred. No.: | 2,15e-291 | Length: | 4679 |
| Score: | 2764.50 | Matches: | 496 |
| Percent Similarity: | 89.98% | Conservative: | 43 |
| Best Local Similarity: | 82.80% | Mismatches: | 59 |
| Query Match: | 85.04% | Indels: | 1 |
| DB: | 9 | Gaps: | 1 |

US-09-807-802a-15 (1-599) x US-09-945-681-10 (1-4679)
Qy 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSer 20
Db 2614 ACAGCTCCGGGAAAAAAGAGAGCCGCTAGAGCACTCTCTCTGGAGCCAGACTCTCTCTCG 2673
Qy 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAenPheGlyGlnThrGly 40
Db 2674 GGAACCGGAAAGCGGCGCCAGCAGCTTGCAGAAAAAGATTGAATTTTGTTCAGACTGGA 2733
Qy 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
Db 2734 GAGCGACACTCAGTACTGACCCCGGAGCTCTCGAGCGCCACAGCGAGCCCTCTCTGGT 2793
Qy 61 ValGlyProThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAspAenAsnGluGly 80
Db 2794 CTGGGAACATAACGATGGCTACAGCAGTGGCGCCCAATATGCGAGCAATAACGAGGGC 2853
Qy 81 AlaAspGlyValGlyAenAlaSerGlyAenTyrHisCysAspSerThrTropLeuGlyAsp 100
Db 2854 GCGCAGCGAGTGGGTAAATTCCTCGGAAATTTGGCATTTCCACATGGATGGGCGAC 2913
Qy 101 ArgValIleThrThrSerThrArgThrTropAlaLeuProThrTyrAenAenHisLeuTyr 120
Db 2914 AGAGTCATACACACAGCAGCCGAACTGGGCGCTCTGCCACCTTACCAACACCACTCTAC 2973
Qy 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAenHisTyrPheGlyTyrSer 140
Db 2974 AAACAAATTTCCAGCCCAATCA--GGAGCCTCGAAGCAGCAATCACTACTTTGGCTACAGC 3030
Qy 141 ThrProTropGlyTyrPheAspPheAenArgPheHisCysHisPheSerProArgAspTyr 160
Db 3031 ACCCTTGGGGGTATTTTGACTTCAACAGATTCACCTGCCACTTTTCCACCGTGAATGG 3090
Qy 161 GlnArgLeuIleAenAenAenTropGlyPheArgProLysArgLeuAenPheLysLeuPhe 180
Db 3091 CAAAGACTCATCAACAACTGGGATTCGAGCCCAAGAGAGACTCACTTCAAGCTCTTT 3150
Qy 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAenAsnLeu 200
Db 3151 AACATTCAGTCAAGAGGTCAGCAGATGACGAGTACGAGCAGCAGATTTGCCAATAACCTT 3210
Qy 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
Db 3211 ACCAGCAGGTTTCAAGTGTATTACTGACTCGAGTACCGAGTCCCGTACGCTCTCGGCTCG 3270
Qy 221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGly 240
Db 3271 GCGCATCAGGATGCTCTCCCGCGTTCACAGCAGAGCTCTTCATGGTGCCACAGTATGGA 3330
Qy 241 TyrLeuThrLeuAenAenGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
Db 3331 TACCTCACCTGAAACAAACGAGAGTCAGGCACTAGGAGCTCTTCATTTTACTTGCCTGGAG 3390
Qy 261 TyrPheProSerGlnMetLeuArgThrGlyAenAenPheThrPheSerTyrThrPheGlu 280
Db 3391 TACTTCTCTCAGATGCTGCGTACCGGAACAACTTTTACCTTCACTACACTTTTGGAG 3450
Qy 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
Db 3451 GAGCTTCTTTCCACAGCAGCTACGCTCAGCAGCAGAGTCTGGACCGCTCTCATGAATCCT 3510
Qy 301 LeuIleAspGlnTyrLeuTyrTyrLeuAenArgThrGlnAenGlnSerGlySerAlaGln 320
Db 3511 CTCATCGACCACTACCTGTATTACTTGGCAGCAACAACTCCCAAGTGGAAACCAACCCACG 3570
Qy 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAen 340
Db 3571 CAGTCAAGGCTTCAAGTTTTCTCAGGCGGAGCGAGTGACATTTGGGACCACTTAGGAAC 3630
Qy 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAenAsn 360
Db 3631 TGGCTTCTCGACCTCTGTACCGCCAGCAGCGAGTATCAAGACATCTGCGGATAACAAC 3690

QY 361 AasnSerAasnPheThrTrpThrGlyAlaSerLysTyxAasnLeuAasnGlyArgGluSerIle 380
DB :
DB 3691 AACAGTGAATACTCGTGAGCTGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTG 3750

QY 381 ILAasnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet 400
DB :
DB 3751 GTGAATCCGGGCCCGGCATGGCAAGCCACACAGGACGATGAAGAAAAGTTTTTTCTCAG 3810

QY 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAasnThrAlaLeuAspAasn 420
DB :
DB 3811 AGCGGGGTTCATCTTTGGAAAGCAAGGCTCAGAGAAACAATGTGGNCATGAAAAAG 3870

QY 421 ValMetIleThrAspGluGluGluIleLysAlaThrAasnProValAlaThrGluArGPhe 440
DB :
DB 3871 GTCATGATTACAGCAGAAGAGGAAATCAGGACAACCAATCCCCTGGCTACGGAGCAGTAT 3930

QY 441 GlyThrValAlaValAasnPheGlnSerSerSerThrAaspProAlaThrGlyAaspValHis 460
DB :
DB 3931 GGTTCTGTATCTACCAACCTCCAGAGAGGCAACAGACAACAGCAGCTCCGAGATGTCAAC 3990

QY 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTrpLeuGlnGlyPro 480
DB :
DB 3991 ACACAAGCGCTTCTCCAGGCGATGGTCTGGCAGCAGACAGATGTGTACCTTCAGGGGCCCC 4050

QY 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
DB :
DB 4051 ATCTGGCAAGAATTCACACACGGACGGACATTTTACCCCTCTCCCTCATGGTGGGA 4110

QY 501 PheGlyLeuLysAasnProProGlnIleLeuIleLysAasnThrProValProAlaAasn 520
DB :
DB 4111 TTCGGACTTAACACACCCCTCTCCACAGATTCTCATCAAGAACACCCCGGTACCTGGCAAT 4170

QY 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
DB :
DB 4171 CCTTCGACCACCTTCAGTGGCGGCAAGTTTGTCTTCATCACACAGTAGTACTCCAGGGA 4230

QY 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAasnSerLysArqTrpAasnPro 560
DB :
DB 4231 CAGETCAGCTGGAGATCGAGTGGAGCTGCAGAGAGGAAACAGCAAACGCTGGATCCC 4290

QY 561 GluValGlnTyrThrSerAasnTyrAlaLysSerAlaAasnValAaspPheThrValAaspAasn 580
DB :
DB 4291 GNAATTCACTACACTTCCAATAACAAGTCTGTTAATGTGGACTTTACTGTGGACACT 4350

QY 581 AasnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
DB :
DB 4351 AATGGCGTGATTTCAGAGCCCTCGCCCATTTGGCCACCAGATACCTTGACTCGTAATCTG 4407

RESULT 9
US-10-136-819-6
; Sequence 6, Application US/10136819
; Publication No. US20030166593A1
; GENERAL INFORMATION:
; APPLICANT: Hoshijima, Masahiko
; TITLE OF INVENTION: No. US20030166593A1-viral vesicle vector for cardiac specific gene
; FILE REFERENCE: 6627-PAL198
; CURRENT APPLICATION NUMBER: US/10/136,819
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: 60/287,423
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 4679
; TYPE: DNA
; ORGANISM: adeno-associated virus 2
US-10-136-819-6

Alignment Scores:
Pred. No.: 2,15e-291 Length: 4679
Score: 2764.50 Matches: 496


```
Db 3511 CTCATCGACCACTACTGCTATTACTTGTGACAGAACAAACACTCCCAAGTGGACCAACCACG 3570
Qy 321 AenLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAen 340
Db 3571 CAGTCAAGGCTTCAGTTCCTCAGGCGGAGCGAGTGCATCTCGGAGCAACGAGTCTAGGAAC 3630
Qy 341 TrpLeuProGlyProCysTrpArgGlnGlnArgValSerLysThrLysThrAspAenAen 360
Db 3631 TGGCTTCCTGGACCTCTGTTACCGCCAGCGAGGATCAACAGACATCTCGCGATCAACAC 3690
Qy 361 AenSerAsnPheThrTrpThrGlyAlaSerLysThrAsnLeuAenGlyArgGluSerIle 380
Db 3691 NACAGTGAATCTCGTGGACTGGAGCTACCAAGTACCACCTCAATGGCAGACAGACTCTCTG 3750
Qy 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet 400
Db 3751 GTGAATCCGGCGCGCCCATGCAAGCCCAACAGGACGATGAAGAAAGTTTTTCTCTCAG 3810
Qy 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAen 420
Db 3811 AGCGGGGTTCTCATCTTTGGGAAGCAAGGCTCAGAGAAACAAATGTGGACATTTGGAAG 3870
Qy 421 ValMetIleThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
Db 3871 GTCATGATTACAGACGAAGAGGAATCAGACAAACCAATCCCGTGGCTACGGAGCAGTAT 3930
Qy 441 GlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHis 460
Db 3931 GGTTCTGTATCTACCAACCTCCAGAGAGGCAACAGACAGCTACCCGACAGATGTCAAC 3990
Qy 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTrpLeuGlnGlyPro 480
Db 3991 ACAGAAGGGCTTTCCAGCATGTGTTCCAGCATGTGTTCCAGCATGTGTTCCAGGGGCC 4050
Qy 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
Db 4051 ATCTGGGCAAGATTCCACACACGGACGGACATTTTCAACCCCTCTCCCTCATGGGTGA 4110
Qy 501 PheGlyLeuLysAsnProProGlnIleLeuLysAsnThrProValProAlaAen 520
Db 4111 TTCCGACTTAAACACCTCTCTCCACAGATTCTCATCAAGAACACCCCGTACTCTCGAAT 4170
Qy 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTrpSerThrGly 540
Db 4171 CCTTCAGCACCTTCAGTGGCGGCAAGTTTGTCTCTCATCACAGTACTCCAGGGA 4230
Qy 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAenPro 560
Db 4231 CAGGTCAGCGTGGAGATCGAGTGGAGCTGCAGAGGAGAAACAGCAACCGTGGAAATCCC 4290
Qy 561 GluValGlnTrpThrSerAsnTrpAlaLysSerAlaAsnValAspPheThrValAspAen 580
Db 4291 GAAATTCAGTACACTTCCCAACTACAAAGCTGTGTAATGTGGACTTTTACTGTGGACAT 4350
Qy 581 AsnGlyLeuTrpThrGluProArgProIleGlyThrArgTrpLeuThrArgProLeu 599
Db 4351 AATGGCGTGTATTCAGAGCCCTCCGCCATTTGGCACCAAGATACCTGACTCGTAATCTG 4407

RESULT 11
US-10-291-583-120
; Sequence 120, Application US/10291583
; Publication No US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alvira, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; FILE REFERENCE: US-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; PRIOR FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
```

```
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 120
; LENGTH: 2205
; TYPE: DNA
; ORGANISM: adeno-associated virus serotype, clone A3.1.vp1
US-10-291-583-120

Alignment Scores:
Pred. No.: 1,42e-291 Length: 2205
Score: 2761.50 Matches: 497
Percent Similarity: 89.98% Conservative: 42
Best Local Similarity: 82.97% Mismatches: 59
Query Match: 84.94% Indels: 1
DB: 13 Gaps: 1

US-09-807-802A-15 (1-599) x US-10-291-583-120 (1-2205)
Qy 1 ThrAlaProGlyLysLysAspProValGlnSerProGlnGluProAspSerSer 20
Db 412 ACGGCTCTCGGAAAGAGAGACCTATAGAGCAGTCTCTCGCAGAACCGGACTTCTCTCG 471
Qy 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAenPheGlyGlnThrGly 40
Db 472 GGCATCGCAATCAGCGCAGCAGCCGCTAAGAAAGACTCAATTTTGGTCAGACTCGC 531
Qy 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAa 60
Db 532 GACACAGAGTCAGTCCCGCAGACCTCAACCAATCGAGAAACCCCGCAGCCCCCTCTCGT 591
Qy 61 ValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAenAsnGly 80
Db 592 GTGGGATCTAATACATGGCTTCAGGCGTGGGCGACCAATGGCAGACAATAACGAGGC 651
Qy 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
Db 652 GCGCAGCGAGTGGGTAAATTCCTCGGGAATTTGGCATTCGATTCACATGATGGCGCAC 711
Qy 101 ArgValIleThrThrSerThrArgTrpAlaLeuProThrTrpAsnHisLeuTrp 120
Db 712 AGAGTTATCACCCAGCAGCAAGAACCTTGGGCCCTCCACCCTACATANTACCTCTAC 771
Qy 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAenHisLeuTrpSer 140
Db 772 AAGCAATCTCCAGCGAATCG--GGAGCCACCAACGACCACTACTTCGGCTACAGC 828
Qy 141 ThrProTrpGlyTrpPheAspPheAsnArgPheHisCysHisPheSerProArgSer 160
Db 829 ACCCCCTCGGGGTATTTTGGACTTTTAAACAGATTCCTCTCTCTCCACCGTACTGG 888
Qy 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAenPheLysLeuPhe 180
Db 889 CAGCGACTCATCACAACACACTGGGGATTTAGACCCAGAACCACTCAATTTCAAGCTCTC 948
Qy 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAenAenLeu 200
Db 949 AACATCCAGTCAAGGAGGTCAAGCAGAAATGATGGAACCCAGCACCATCGCCAAATACCTT 1008
Qy 201 ThrSerThrValGlnValPheSerAspSerGluTrpGlnLeuProTrpTrpValLeuGlySer 220
Db 1009 ACCAGCAGCGTGCAGGTCTTCACAGACTCTGAGTACAGCTGCCCTACCTCTCGGTTCC 1068
Qy 221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTrpGly 240
Db 1069 GTCACACAGGGTGTCTTCGCGCGTTCAGCAGCAGACGCTTCATGATCTCTCAGTACGCG 1128
Qy 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
```

Db 1129 TACTTGACTCTGAACAATAGCGAGCGTAGGACGCTTCTTATTCTACTGCTCTAGAG 1188
Qy 261 TyrPheProSerGlnMetLeuArgThrGlyAenAenPheThrPheSerTyrThrPheGlu 280
Db 1189 TATTTTCCTCTCAGATCTGAGGACGGGAACAACATTCACCTTCAGTACACATTTTGA 1248
Qy 281 GluValProPheHisSerSerTyrTAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
Db 1249 GACGTGCTTTTCCACAGCAGCTACGCGCACAGCAGCAGCTCTGGATCGGCTGATGAATCCT 1308
Qy 301 LeuLeuAspGlnTyrLeuTyrTyrLeuAenArgThrGlnAenGlnSerGlySerAlaGln 320
Db 1309 CTCATTGACCACTGATCTTATCTGAGCAAACTCAGGTACAAAGTGGAAACACGCGAG 1368
Qy 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
Db 1369 CAATCGACACTGCAGTTACGCAAGCTGGGCTAGCTCCATGGCTCAGCAGCGCCAAAC 1428
Qy 341 TrpLeuProGlyProCysTyrArgGlnArgValSerLysThrLysThrAspAsnAsn 360
Db 1429 TGGCTACCGGACCCAGCTACCGACAGCGAGATGTCTAAGACGGCTAATGACACAAC 1488
Qy 361 AsnSerAsnPheThrTyrThrGlyAlaSerLysTyrAsnLeuAenGlyArgGluSerIle 380
Db 1489 AACAGTGAATTTGCTTGACTGCAGCCCAACAAATATTACCTGAATGGAGAAATTTCTGT 1548
Qy 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet 400
Db 1549 GTCAATCCCGGGCCCCCAATGCCAGTCAAGAGACGATGAGAAAGATATTTCCCAATG 1608
Qy 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
Db 1609 CACGGAATCTCATCTTTGGAAAACAAGGCACAGCACTACCAATGTGGACATTGAATCA 1668
Qy 421 ValMetIleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
Db 1669 GTGCTTATTCAGACCAAGAAATCAAGAACTAATCTCTGTGGCTACAGAAACAATAC 1728
Qy 441 GlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHis 460
Db 1729 GGACNGTGGCCACCAACATCAGAGTCAGAACACCACTGCTTCATGGAGTGTGGAC 1788
Qy 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro 480
Db 1789 AGCCAGGGAATCTTACCTGGAATGTGTGGCAGCAGCCGATGCTATCTTCAAGTCCC 1848
Qy 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
Db 1849 ATTTGGGCCAAACTCTTCACAGCAGCAGCACTTTTCATCTCTCCGCTCATGGAGGC 1908
Qy 501 PheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsn 520
Db 1909 TTTGACCTGAAACACCTCTCTCCCAAGATCTGATCAAAAACACACCTGTGCCAGCGAAT 1968
Qy 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
Db 1969 CCGCGACCACTTTCACTCTCTGGAAAGTTTGTCTGTTTCATTACCCAGTATTCACCGGA 2028
Qy 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAenSerLysArgTyrAsnPro 560
Db 2029 CAGGTGAGCTGGAATAGATGGAGCTGCAGAAAGAAACACGAAACGCTGGAAACCCA 2088
Qy 561 GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsn 580
Db 2089 GAAATTCAGTACCTTCCAACTACAAACAGTCGGTGAATGTGGATTTACCGTGGACGCCA 2148
Qy 581 AsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db 2149 AACGGTGTATTCTGAAACCCCGCCCTATTGGCACTCGTTACCTTACCCGGAACCTTG 2205

RESULT 12

US-10-291-583-56

; Sequence 56, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alvirra, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; FILE REFERENCE: UPN-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56
; LENGTH: 3122
; TYPE: DNA
; ORGANISM: new AAV serotype, clone A3.7
US-10-291-583-56

Alignment Scores:

| | | | |
|------------------------|-----------|---------------|------|
| Pred. No.: | 1,83e-290 | Length: | 3122 |
| Score: | 2753.50 | Matches: | 496 |
| Percent Similarity: | 89.82% | Conservative: | 42 |
| Best Local Similarity: | 82.80% | Mismatches: | 60 |
| Query Match: | 84.70% | Indels: | 1 |
| DB: | 13 | Gaps: | 1 |

US-09-807-802A-15 (1-599) x US-10-291-583-56 (1-3122)

| | | | |
|----|------|--|------|
| Qy | 1 | ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSer | 20 |
| Db | 1258 | ACGGCTCTGGAAAAAGAGACCTATAGACAGCTCTCTGCAGAACCGGACTCTTCCTCG | 1317 |
| Qy | 21 | GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAenPheGlyGlnThrGly | 40 |
| Db | 1318 | GGCATCGGCAATCAGCCAGCAGCCGCTAAGAAAAGACTCAATTTTGGTCAGACTGGC | 1377 |
| Qy | 41 | AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla | 60 |
| Db | 1378 | GACACAGATCAGTCCCGACACCTCAACCAATCGGAGAACCCCGCAGCCCTCTCGT | 1437 |
| Qy | 61 | ValGlyProThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGluGly | 80 |
| Db | 1438 | GTGGATCTAATCAATGGCTTCAGCGGTGGGGCACCATAATGGCAGCAATAACGAAGGC | 1497 |
| Qy | 81 | AlaAspGlyValGlyAenAlaSerGlyAenTyrHisCysLeuSerThrTyrLeuGlyAsp | 100 |
| Db | 1498 | GCCGACGAGTGGGTAAATCTCTCGGAAATGGCATTCGATTCACATGGATGGCGAC | 1557 |
| Qy | 101 | ArgValIleThrThrSerThrArgThrTyrAlaLeuProThrTyrAsnAsnHisLeuTyr | 120 |
| Db | 1558 | AGAGTTATCACCACAGCAGCAGCAAGACCTGGGCGCTCCACCTACATATATCGCTCTAC | 1617 |
| Qy | 121 | LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer | 140 |
| Db | 1618 | AAGCAATCTCCAGCGAATCG--GGAGCCACCAACGACCAACCACTACTCTCGGTACAGC | 1674 |
| Qy | 141 | ThrProTyrGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTyr | 160 |
| Db | 1675 | ACCCCTGGGGGTATTTTGACTTTAACAAGATTCACATGTCCTCTCTCACCACGAGACTGG | 1734 |
| Qy | 161 | GlnArgLeuIleAenAsnAsnTyrGlyPheArgProLysArgLeuAenPheLysLeuPhe | 180 |
| Db | 1735 | CAGGACTCATCAACAACAACTGGGGATTTAGACCCCAAGAACTCAATTTTCAGCTCTTC | 1794 |

Qy 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAenHisPheGlyTyrSer 140
Db AAGCAAAATCTCCAGCGAATCG--GGAGCCACCAAGAGCAACCACTACTCTCGGCTACAGC 1665
Qy 141 ThrProTTPGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
Db ACCCCCTGGGGGTATTTTGAATTTAAACAGATTCACATGCTCATCTCCACCAAGTGACTGG 1725
Qy 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnAsnLeuPhe 180
Db CAGGACTCATCAACCAACACTGGGGATTTAGACCCAGAAACTCAATTCAGACTCTTC 1785
Qy 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeu 200
Db AACATCCAAAGTCAGAGAGTCACGCAAGATGATGGAACCAAGCAACATCGCAATAACCTT 1845
Qy 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
Db ACCAGCAGGGTGCAGGTCTTCACAGACTCTGAGTACCAAGCTGCCCTCAGTCTCGGTTCG 1905
Qy 221 AlaHisGlnGlyCysLeuProPheProPheProAlaAspValPheMetIleProGlnTyrGly 240
Db GCTCACAGGGCTGCTTCGCGCTCCAGCAGAGCTCTTCATGATTCCTCAGTACGAGC 1965
Qy 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
Db TACTTGACTCTGAACAATGGCAGCAGCAAGCGGTAGGACGCTTCTCATTCTACTGTCTAG 2025
Qy 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnAsnPheThrPheSerTyrThrPheGlu 280
Db TATTTTCCTCTCAGATGCTGAGGACGGGAACAACATTCACCTTCAGTACACTTTTGA 2085
Qy 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAsnArgLeuMetAsnPro 300
Db GAGGTGCTTCCACAGCAGCTACCGGACAGCAGCTGCTGATCGCTGATGATCTCT 2145
Qy 301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
Db CTCATTGACCACTGCTGATTTACCTGAGCAAACTCAGGGGTACAAAGTGGAAACACGAG 2205
Qy 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
Db CAATCGAGACTCAGTTACGACCACTGGCGCTAGCTCCATGGCTCAGCAGCCCAAAAC 2265
Qy 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
Db TGGCTACCGGACCCAGCTACCGACAGCAGCGAATGCTAGACGGCTAATGACAAAC 2325
Qy 361 AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 380
Db AACAGTGAATTTGCTTGGACTCGACCCACCAATATTTACCTGAATGGAAATTTCTCTG 2385
Qy 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPheProMet 400
Db GTCATATCCCGGGCCCCCAATGGCCAGTACAAAGGACGATGAGGAAAGATTTTCCCAATG 2445
Qy 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
Db CACGGAAATCTCATCTTTGGAAAAACAGGCACAGGACTACCAATGTGGACATGAATCA 2505
Qy 421 ValMetIleThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
Db GTGCTTATTACAGACGAAGAAATCAGAAACAATTAATCTCTGTGGCTACAGAAATAC 2565
Qy 441 GlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHis 460
Db GGACAGGTGGCCCAACCATCAGAGTCAGGACCAACCAAGCTTCTTATGAGAGTGGAC 2625
Qy 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro 480
Db ACCCAGGGAATCTTACCTGGATGGTGTGGCAGGCGGATGTCTATCTTCAAGGTCC 2685

Qy 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
Db ATTTGGCCAAAATCTCTCACAGGACGACACTTTTCATCTCTCCGCTCATGGAGGC 2745
Qy 501 PheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsn 520
Db TTTGGACTGAACACCTCTCTCCAGATCTGATCTAAACACACACTGTGCCAGCGAAT 2805
Qy 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
Db CCGCGCACCACTTTCACCTCTCGGAAGTTTGTCTGTTTCATTCACCAAGTATTCACCGGA 2865
Qy 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPro 560
Db CAGGTACGCTGGAAATAGAGTGGAGCTCGAGAAAGAAACAGCAACCGCTGGAACCCA 2925
Qy 561 GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsn 580
Db GAAATTCAGTACCTCAACTACACAGTGGTGAATGTGGAGTTTACCGTGGACCGCA 2985
Qy 581 AsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db AACGGTGTATTCTGAACCCCGCCCTATTGGCACTGTTACCTTACCTTACCCGGAACCTG 3042
RESULT 14
US-10-291-583-55
; Sequence 55, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alwira, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; FILE REFERENCE: UPN-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55
; TYPE: DNA
; ORGANISM: new AAV serotype, clone A3.5
US-10-291-583-55
Alignment Scores:
Pred. No.: 5e-290 Length: 3113
Score: 2749.50 Matches: 495
Percent Similarity: 89.82% Conservative: 43
Best Local Similarity: 82.64% Mismatches: 60
Query Match: 84.57% Indels: 1
DB: 13 Gaps: 1
US-09-807-802A-15 (1-599) x US-10-291-583-55 (1-3113)
Qy 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSer 20
Db 1249 ACGGCTCTGGAAGAAAGAGACCTATAGCAGCTCTCTCGAGAACCGGACTTCTCTCG 1308
Qy 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGly 40
Db 1309 GGCATCGCAATCAGGCCAGCAGCCGCTAAGAAAAGACTCAATTTTGGTGCAGACTGCG 1368
Qy 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
DB: 1

1369 GACACAGAGTCAGTCCAGACCTCAACCAATCGGAGAACCCCGCAGCCCTCTGGT 1428
Qy ValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnGly 80
1429 GTGGGATCTAATACCAATGCTTCAGCGGTGGGGACCAATGGCAGACAATAACGAGGC 1488
Qy AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
1489 GCCGACGGAGTGGGTAATTCCTCGGGAATGGCATGGCATTCACATGGATGGCGAC 1548
Qy ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTyAsnAsnHisLeuTy 120
1549 AGAGTTATCACCCAGCAGCAAGAACCTGGGCGCTCCCGACCTCAACAATCACCTTAC 1608
Qy LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAsnHisTyPheGlyTySer 140
1609 AAGCAATCTCCAGCGAATCG--GGAGCCACCAACGACCACTACTTCGGCTACAGC 1665
Qy ThrProTrpGlyTyPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
1666 ACCCTCGGGGTATTTGACTTTACAGATTCACATTCCTGCTGCTTCCACAGCTGCTGG 1725
Qy GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
1726 CAGCGACTCATCAATAACAACTGGGATTTAGACCCCAAGAACTCAATTTCAAGCTCTC 1785
Qy AsnIleGlnValLysGluValThrAsnAspGlyValThrIleAlaAsnLeu 200
1786 AACATCCAAGTCAAGAGGTACGCGAGATGTATGGAAACCAACCACTCCCAATACCTT 1845
Qy ThrSerThrValGlnValPheSerAspSerGluTyGlnLeuProTyValLeuGlySer 220
1846 ACCAGCAGCGTGCAGGTCTTCACAGACTCTGAGTACAGCTGCCCTGCTCTCGGTTG 1905
Qy AlaHisGlnGlyCysLeuProPropheProAlaAspValPheMetIleProGlnTyGly 240
1906 GCTCACCAGGGTGCCTTCGCGCTCCAGCAGAGCTCTCATGATTCCTCAGTACGGC 1965
Qy TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyCysLeuGlu 260
1966 TACTTGACTCTGAACATGGCAGCCAGCGGTAGGACGTTCTTCATCTACTGTCTAGAG 2025
Qy TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyThrPheGlu 280
2026 TATTTCCCTCTCAGATGCTGAGAGCGGGAACAATTCACCTTCAGCTACACTTTTGA 2085
Qy GluValProPheHisSerSerTyAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
2086 GAGGTGCTTTCCACAGCAGCTACGGCGCACAGCCAGAGTCTGGATCGGCTGATGATCT 2145
Qy LeuIleAspGlnTyLeuTyTyTyLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
2146 CTCAATGACAGTACCTGTATTAACCTGAGCAAACTCAGGTTCAAGTGAACACGCGAG 2205
Qy AsnLysAspLeuLeuPheSerArgLysSerProAlaGlyMetSerValGlnProLysAsn 340
2206 CAATCGAGACTGCAGTTCAACCAAGCTGGCGCTAGCTCCAGTCCAGCAGGCGCAAAAC 2265
Qy TrpLeuProGlyProCysTyTyArgGlnArgValSerLysThrLysThrAspAsnAsn 360
2266 TGGCTACCGGGACCCAGCTACCGCAGCAGCGAATGTCTAAGACGCGTAATGACACAAC 2325
Qy AsnSerAsnPheThrTrpThrGlyAlaSerLysTyAsnLeuAsnGlyArgGluSerIle 380
2326 AACAGTGAATTTGCTGGATCTGCGCCGCCCAATATTAACCGAATGGAGAAATTTCTTG 2385
Qy IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPheProMet 400
2386 GTCAATCCCGGGCCCCCAATGGCCAGTCAACAGGACGATGCGGAAAGTATTTCCCATG 2445
Qy SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
2446 CACGGAATCTCATCTTTGGAAAAACAAGGCACAGGAACTACCAATGTGGCATTTGAATCA 2505

Qy ValMetIleThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
Db GTGCTTATTCAGACGAGGAGAAATCAGAACGATCTATCTCTGGCTACAGACATATAC 2565
Qy GlyThrValAlaValAsnAsnGlnSerSerThrAspProAlaThrGlyAspValHis 460
Db GGACAGGTGTCACCAACCGTCAGAGTCAGAACACACAGCTTCTCTATGGAGTGTGAC 2625
Qy AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyLeuGlnGlyPro 480
Db AGCCAGGGAATCTTACCTGGAAATGCTGGCAGACCGCGATGTCTATCTTCAAGGTCCC 2685
Qy IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
Db ATTTGGGCCAAACTCTTCACAGGACGACACTTTCATCTCTCCGCTCATGGAGGC 2745
Qy PheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsn 520
Db TTTGGACTGAAACACCTCTCCAGATCTGATCAAAAACACACCTGTGCCAGCGAAT 2805
Qy ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTySerThrGly 540
Db CCCCGACCACTTTCATCTCTGGAAGTTCTCTGTTTCATTCACCACTTCCACCGGA 2865
Qy GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPro 560
Db CAGTTCAGCTGGAAATAGTGGAGCTGGAGAAAGAAACAGACAAACGCTGGAACCG 2925
Qy GluValGlnTyThrSerAsnTyAlaLysSerAlaAsnValAspPheThrValAspAsn 580
Db GAAATTCAGTACACTCCCACTACAAACAGTCCGTAATGTGGAGTTTACCGTGGACGCA 2985
Qy AsnGlyLeuTyThrGluProArgProIleGlyThrArgTyLeuThrArgProLeu 599
Db AACGTGTGTTTATCTGAACCCCGCTTATGGCACTCTGTTACCTTACCGGAACTTG 3042

RESULT 15

US-10-291-583-57
Sequence 57, Application US/10291583
Publication No. US20030138772A1
GENERAL INFORMATION:
APPLICANT: Gao, Guangping
APPLICANT: Wilson, James M.
APPLICANT: Alvira, Mauricio
TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
TITLE OF INVENTION: Sequences and Isolating No. US20030138772A1el Sequences Identif
FILE REFERENCE: UPN-0273505A
CURRENT APPLICATION NUMBER: US/10/291,583
PRIOR FILING DATE: 2002-11-12
PRIOR APPLICATION NUMBER: US 60/350,607
PRIOR FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 60/341,117
PRIOR FILING DATE: 2001-12-17
PRIOR APPLICATION NUMBER: US 60/377,066
PRIOR FILING DATE: 2002-05-01
PRIOR APPLICATION NUMBER: US 60/386,675
PRIOR FILING DATE: 2002-06-05
NUMBER OF SEQ ID NOS: 120
SOFTWARE: PatentIn version 3.1
SEQ ID NO 57
LENGTH: 3123
TYPE: DNA
ORGANISM: new AAV serotype, clone A3.3
US-10-291-583-57

Alignment Scores:
Pred. No.: 1,71e-288 Length: 3123
Score: 2735.50 Matches: 493
Percent Similarity: 89.48% Conservative: 43
Best Local Similarity: 82.30% Mismatches: 62
Query Match: 84.14% Indels: 1
DB: 13 Gaps: 1

US-09-807-802A-15 (1-599) x US-10-291-583-57 (1-3123)

```
Qy 1 ThrAlaProGlyLysLeuValGluGlnSerProGlnGluProAspSerSerSer 20
Db 1249 ACGGCTCTCTGGAAAGAGACCTATAGAGCAGTCTCTCGCAGAACCGGACTTCTCTCG 1308
Qy 21 GlyLeuGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGly 40
Db 1309 GGCATCGGCAGAAATCAGGCAGCAGCCGCTAAGAAAAGACTCAATTTTGGTCAGACTGGC 1368
Qy 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
Db 1369 GACACAGAGTCAGTCCCGAGGCCCTCAACCAATCGGAGAACCCCGCAGGCCCTCTGGT 1428
Qy 61 ValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGly 80
Db 1429 GTGGGATCTAATACATTCGCTTCAGCGGTGGGACCAATGGCAGACATTAACGAGGC 1488
Qy 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
Db 1489 GCGGACGGAGTGGGTAAATTCCTCGGGAATTTGGCATTCGGATTCCACATGATGGCGAC 1548
Qy 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrThrAsnAsnHisLeuTyr 120
Db 1549 AGAGTTATCACCCACAGCACAGAACCTGGGCGCTCCCGCCTACCAATTAATCACCTCTAC 1608
Qy 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer 140
Db 1609 AAGCAAAATCTCCAGCGAATCG---GGAGCCACCAACGACCACTACTTCGGCTACAGC 1665
Qy 141 ThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
Db 1666 ACCCCCTGGGGTATTTTGACTTTAAAGATTCACACTGTCACTTCTCACACAGTACTGG 1725
Qy 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheIleLeuPhe 180
Db 1726 CAGCGACTCATCAACAACTGGGGATTTAGACCCAGAACTCAATTTCAAGCTCTTC 1785
Qy 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeu 200
Db 1786 AACATCCAACTCAGAGAGTTCAGCGAATGATGGAACCAAGCACCATTCCGCATACCTT 1845
Qy 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
Db 1846 ACCAGCGGGTGCAGGTCTTACAGACTCTGAGTACAGCTGCCCTACGCTCTCGTTTCG 1905
Qy 221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGly 240
Db 1906 GCTCACCAAGGGCTGCTTCCGCGCTCCAGCAGACGCTTCTCATGATTCCTCAGTACGGC 1965
Qy 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
Db 1966 TACTTGACTCTGAACAATGGCAGCCAGCGGTAGACGTTCTTCATTTCTACTGCTAGAG 2025
Qy 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
Db 2026 TATTTTCCCTCTCAGATGCTGAGGACGGGAAACAACTTCACCTTCAGCTACACTTTTGA 2085
Qy 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
Db 2086 GACGTGCTCTTCCACAGCAGCTACCGGCACAGCCAGAGTCTGGATCGGCTGATGAATCCT 2145
Qy 301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
Db 2146 CTCATTGACCACTGCTATTTACTCTGAGCAAACTCAGGGTACAGTGGGACACACGCGAG 2205
Qy 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
Db 2206 CAATCGAGACTCGAGTTACGCCAAGCTGGGCTAGCTCCATGGCTCAGCGGCCCAAAAC 2265
Qy 341 TrpLeuProGlyProCysThrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
```

```
Db 2266 TGGCTACCCGGACCCAGCTACCACAGCAGCGAATGTCTAAGACGGCTAATGACAAACAC 2325
Qy 361 AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 380
Db 2326 AACAGTGAATTTGCTTGGACTGCAGCCACCAATATTTACCTGAATGGAGAAATTCCTG 2385
Qy 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet 400
Db 2386 GTCAATCCCGGGCCCCAGTGGCCAGTCACAAGCAGCATGAGGAAAAGTATTTCCCATG 2445
Qy 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
Db 2446 CACGAAATCTCATCTTTGGAAACACAGGCACAGAACTACCAATGTGGACATTTGAATCA 2505
Qy 421 ValMetIleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
Db 2506 GTGCTTATTACAGACGGAAGAAATCAGAACAACATAATCTCTGTGGCTACAGAACAAATAC 2565
Qy 441 GlyThrValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHis 460
Db 2566 GGACAGGTTGCCACCAACCATCAGAGTCAGAGTCAGAACACCAACAGCTTCTATGGAAGTGTGAC 2625
Qy 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro 480
Db 2626 AGCAGGAAATCTTACCTGGAATGGTGTGCAGACCCGCGATGTCTATCTTCAAGTCCC 2685
Qy 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
Db 2686 ATTTGGGCCAAAACTCCTCACACGGACGGACACTTTTCATCTCTCGCTCATGGGAGGC 2745
Qy 501 PheGlyLeuLysAsnProProProGlnIleLeuIleLysAsnThrProValProAlaAsn 520
Db 2746 TTTGGACTGAACACCCCTCTCCCGAGATCTGATCAAAAACACACCTGTGCCAGCGAAT 2805
Qy 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
Db 2806 CCGCGACCACTTTTCACTCTCTGGAAAGTTTGTCTGTTCTATTACCCAGTATTCACCTGA 2865
Qy 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPro 560
Db 2866 CAGGTCAAGCTGGAAATAGAGTGGAGCTGCAGAAAGAAACAGCAAAACGCTGGAACCCA 2925
Qy 561 GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsn 580
Db 2926 GAAATTCAGTACACCTCCAACTACAACAGTCGGTGAATGTGGAGTTTACCCTGCGACGCA 2985
Qy 581 AsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db 2986 AACGGTGTATTCTGAAACCCCGCCCTATTGGCACTCTGTACCTTACCCGGAACCTTG 3042
```

Search completed: February 15, 2004, 02:06:20

Job time : 695.336 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 14, 2004, 16:55:45 ; Search time 496.857 Seconds
(without alignments)
3959.005 Million cell updates/sec

Title: US-09-807-802A-17

Perfect score: 2906

Sequence: 1 MASGGGAPMADNNEGADGVG.....NNGLYTEPRPTGTRYLTRPL 534

Scoring table: BLOSUM62

| | |
|-------------|-------------|
| Xgapop 10.0 | Xgapext 0.5 |
| Ygapop 10.0 | Ygapext 0.5 |
| Fgapop 6.0 | Fgapext 7.0 |
| Delop 6.0 | Delext 7.0 |

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-O=/cgn2_1/USPTO_spool/US09807802/runat_11022004_175610_15987/app_query.fasta_1.2389
-DB=Published Applications NA -QWMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOFCCL=0 -LOOPEXT=0 -UNIT8=bits -START=1 -END=-1 -MAIRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USRP=US09807802 @cgn 1 1100 @runat_11022004_175610_15987
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/ECTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result | Query | Match | Length | ID | Description |
|--------|-------|-------|--------|----|-------------|
| No. | Score | | | | |

| | | | | | | |
|----|--------|-------|------|----|-------------------|-------------------|
| 1 | 2906 | 100.0 | 4718 | 13 | US-10-291-583-6 | Sequence 6, Appl |
| 2 | 2906 | 100.0 | 7447 | 13 | US-10-216-870-11 | Sequence 11, Appl |
| 3 | 2536.5 | 87.3 | 3075 | 13 | US-10-291-583-26 | Sequence 26, Appl |
| 4 | 2528 | 87.0 | 4726 | 13 | US-10-291-583-8 | Sequence 8, Appl |
| 5 | 2520.5 | 86.7 | 3142 | 13 | US-10-291-583-25 | Sequence 25, Appl |
| 6 | 2518 | 86.6 | 8179 | 15 | US-10-205-942-5 | Sequence 5, Appl |
| 7 | 2481.5 | 85.4 | 4679 | 9 | US-09-804-898-1 | Sequence 1, Appl |
| 8 | 2481.5 | 85.4 | 4679 | 9 | US-09-945-681-10 | Sequence 10, Appl |
| 9 | 2481.5 | 85.4 | 4679 | 14 | US-10-136-819-6 | Sequence 6, Appl |
| 10 | 2481.5 | 85.4 | 4679 | 14 | US-10-038-972A-12 | Sequence 12, Appl |
| 11 | 2457.5 | 84.6 | 2205 | 13 | US-10-291-583-120 | Sequence 120, App |
| 12 | 2451.5 | 84.4 | 3123 | 13 | US-10-291-583-54 | Sequence 54, Appl |
| 13 | 2449.5 | 84.3 | 3122 | 13 | US-10-291-583-56 | Sequence 56, Appl |
| 14 | 2445.5 | 84.2 | 3113 | 13 | US-10-291-583-55 | Sequence 55, Appl |
| 15 | 2441 | 84.0 | 2489 | 13 | US-10-291-583-35 | Sequence 35, Appl |
| 16 | 2439.5 | 83.9 | 4680 | 13 | US-10-375-777-1 | Sequence 1, Appl |
| 17 | 2439.5 | 83.9 | 4680 | 14 | US-10-077-294-1 | Sequence 1, Appl |
| 18 | 2439.5 | 83.9 | 4680 | 14 | US-10-163-886-1 | Sequence 1, Appl |
| 19 | 2439.5 | 83.9 | 4680 | 15 | US-10-263-127-1 | Sequence 1, Appl |
| 20 | 2439 | 83.9 | 3098 | 13 | US-10-291-583-9 | Sequence 9, Appl |
| 21 | 2438.5 | 83.9 | 3123 | 13 | US-10-291-583-57 | Sequence 57, Appl |
| 22 | 2437.5 | 83.9 | 3106 | 13 | US-10-291-583-34 | Sequence 34, Appl |
| 23 | 2436 | 83.8 | 2495 | 13 | US-10-291-583-36 | Sequence 36, Appl |
| 24 | 2436 | 83.8 | 3098 | 13 | US-10-291-583-37 | Sequence 37, Appl |
| 25 | 2436 | 83.8 | 3106 | 13 | US-10-291-583-18 | Sequence 18, Appl |
| 26 | 2434.5 | 83.8 | 3123 | 13 | US-10-291-583-41 | Sequence 41, Appl |
| 27 | 2431.5 | 83.7 | 2504 | 13 | US-10-291-583-33 | Sequence 33, Appl |
| 28 | 2430 | 83.6 | 4721 | 13 | US-10-291-583-1 | Sequence 1, Appl |
| 29 | 2429.5 | 83.6 | 3129 | 13 | US-10-291-583-59 | Sequence 59, Appl |
| 30 | 2428.5 | 83.6 | 3128 | 13 | US-10-291-583-27 | Sequence 27, Appl |
| 31 | 2428.5 | 83.6 | 3197 | 13 | US-10-291-583-29 | Sequence 29, Appl |
| 32 | 2427.5 | 83.5 | 3084 | 13 | US-10-291-583-39 | Sequence 39, Appl |
| 33 | 2423.5 | 83.4 | 4675 | 10 | US-09-782-378A-1 | Sequence 1, Appl |
| 34 | 2423.5 | 83.4 | 4675 | 10 | US-09-782-378A-2 | Sequence 2, Appl |
| 35 | 2423.5 | 83.4 | 4675 | 13 | US-10-291-583-7 | Sequence 7, Appl |
| 36 | 2423.5 | 83.4 | 4675 | 15 | US-10-240-198-1 | Sequence 1, Appl |
| 37 | 2423 | 83.4 | 3098 | 13 | US-10-291-583-10 | Sequence 10, Appl |
| 38 | 2422.5 | 83.4 | 3113 | 13 | US-10-291-583-31 | Sequence 31, Appl |
| 39 | 2422.5 | 83.4 | 3128 | 13 | US-10-291-583-47 | Sequence 47, Appl |
| 40 | 2422 | 83.3 | 3106 | 13 | US-10-291-583-16 | Sequence 16, Appl |
| 41 | 2421.5 | 83.3 | 2501 | 13 | US-10-291-583-30 | Sequence 30, Appl |
| 42 | 2421.5 | 83.3 | 3121 | 13 | US-10-291-583-11 | Sequence 11, Appl |
| 43 | 2419.5 | 83.3 | 3095 | 13 | US-10-291-583-24 | Sequence 24, Appl |
| 44 | 2418.5 | 83.2 | 3095 | 13 | US-10-291-583-23 | Sequence 23, Appl |
| 45 | 2418 | 83.2 | 3102 | 13 | US-10-291-583-17 | Sequence 17, Appl |

ALIGNMENTS

RESULT 1

US-10-291-583-6
; Sequence 6, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alvira, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; TITLE OF INVENTION: Sequences and Isolating No. US20030138772A1
; FILE REFERENCE: US-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; PRIOR FILING DATE: 2002-11-12
; PRIOR FILING DATE: 2002-11-12
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6

; LENGTH: 4718
; TYPE: DNA
; ORGANISM: adeno-associated virus serotype 1
US-10-291-583-6

Alignment Scores:

Prod. No.: 0 Length: 4718
Score: 2906.00 Matches: 534
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 13 Gaps: 0

US-09-807-802a-17 (1-534) x US-10-291-583-6 (1-4718)

```
QY 1 MetAlaSerGlyGlyAlaProMetAlaAspAsnGluGlyAlaAspGlyValGly 20
DB 2829 ATGGCTTCAGCGGTGGCGCACCACCAATGGCAGACAATAACGAGCGCGCAGGATGGGT 2888
QY 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 40
DB 2889 AATGCCCTCAGGAAATGGCAATTCGATTCACATGGCTGGCGCAGAGTCATCACACC 2948
QY 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSer 60
DB 2949 AGCACCGCACCTGGCGCTGGCCACCTACATAACCACTCTACAGCAATCTCCAGT 3008
QY 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
DB 3009 GCTTCAACGGGGCGGCGAGCAACGACCACTACTTCGGCTACAGCACCCCGCTGGGGGTAT 3068
QY 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
DB 3069 TTGGATTCAACAGATTCCACTGGCTTTCACACGCTGACGGCAGGAGCTCATCAAC 3128
QY 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPhelLysLeuPheAsnIleGlnValLys 120
DB 3129 AACAAATGGGGANTCCGGCCCAAGAGACTCAACTTCAAACTCTTCAACATCCAACTCAAGTC 3188
QY 121 GluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeuThrSerThrValGln 140
DB 3189 GAGGTACAGCAGCAATGATGGCGTCACAAACATCGCTAATAACCTTACAGCAGCGTTCAA 3248
QY 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
DB 3249 GTCTTCTCGGACTCGGAGTACCAGCTTCCGTAGCTCTCGGCTCTGGCCACAGGGCTGC 3308
QY 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
DB 3309 CTCCTCCGTTCCCGCGGACGTTTCATGATCCGCAATACGGCTACCTGACGCTCAAC 3368
QY 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
DB 3369 AATGGCAGCAAGCGCGGAGCTTCATCTCTTTACTGCGCTGGAATATTTCCCTCTTCAG 3428
QY 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
DB 3429 ATGCTGAGAACGGGCAACAATTTTACCTTTCAGCTACACCTTTGAGGAAGTGCCTTCCAC 3488
QY 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
DB 3489 AGCAGCTACGGCAGCAGCAGCGCTGGACCGGCTGATGAATCTCTCTCATCGCAATAC 3548
QY 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
DB 3549 CTGTATTACCTGAAACAGAACTCAAAATCAGTCGGGAAGTGCCTCAAAACAGGACTTGTCTG 3608
QY 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
DB 3609 TTTAGCGGTGGGTCTCCAGCTGGCATGTCTGTTCAGGCCCAAAACTGGCTTACTTGNCCC 3668
QY 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
```

```
DB 3669 TGTATCGCAGCAGCGCGTTTCTAAACAAAAACAGACAACACAGCAATTTTACC 3728
QY 301 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr 320
DB 3729 TGGACTGGTCTTCAAAATATAACCTCAATGGCGGTGAATCATCATCAACCCCTGGCACT 3788
QY 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
DB 3789 GCTATGGCTCTACACAAAGACGACGAGCAAGTTCTTCTCCATGAGCGGTGTCTGATTT 3848
QY 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
DB 3849 TTTGGAAAGAGAGCGCGGAGCTTCAAACTGCTATGGCAATGTCTATGATTACAGAC 3908
QY 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluAtgPheGlyThrValAlaVal 380
DB 3909 GAAGAGGAAATTAAGCCACTAACCTGTGGCCACCGAAGAGATTTGGGACCGTGGCAGTC 3968
QY 381 AsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
DB 3969 AATTTCCAGACGACGACGACGACCTCGGACCGAGATGTGCTATGGGAGCATTA 4028
QY 401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 420
DB 4029 CCTGGCATGGTGTGGCAAGATAGACGCTGCTACCTGACGGGTCCCATTTGGGCCCAAAAT 4088
QY 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
DB 4089 CCTCACAGATGGACACTTTTCCACCGCTCTCTCTTATGGGCGGCTTTGGACCTCAAGAAC 4148
QY 441 ProProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460
DB 4149 CGGCTCTCTCAGATCTCTCATCAAAACACGCTGTCTCTCGAATCTCTCGGCGGATTT 4208
QY 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
DB 4209 TCAGCTCAAAAGTTTGTCTTCAATCATCAACCAATCTCCACAGGACAAAGTGTAGTGTGAA 4268
QY 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrThr 500
DB 4269 ATTGAATGGAGCTCGAGAAAGAACAGCAAGCGCTGGAATCCCGAGTGCAGTACACA 4328
QY 501 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 520
DB 4329 TCCAAATATGCAAAATCTGCCAACCTGTGATTTTACTGTGGACAACAATGGACTTTTATACT 4388
QY 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
DB 4389 GAGCCTCGCCCAATGGCACCGCTTACCTTACCCGTCCTCCCTG 4430
```

RESULT 2

US-10-216-870-11
; Sequence 11, Application US/10216870
; Publication No. US20030148506A1
; GENERAL INFORMATION:
; APPLICANT: KOTIN, ROBERT M
; APPLICANT: URABE, MASASHI
; APPLICANT: DING, CHUAN-TIAN
; TITLE OF INVENTION: PRODUCTION OF ADENO-ASSOCIATED VIRUS IN INSECT CELLS
; FILE REFERENCE: 402133
; CURRENT APPLICATION NUMBER: US/10/216,870
; CURRENT FILING DATE: 2002-08-13
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 11
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-216-870-11

Alignment Scores:

Pred. No.: 0 Length: 7447
Score: 2906.00 Matches: 534
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 13 Gaps: 0

US-09-807-802A-17 (1-534) x US-10-216-870-11 (1-7447)

Qy 1 MetAlaSerGlyGlyValAlaProMetAlaAspAenAenGluGlyAlaAspGlyValGly 20
Db ATGGCTTCAGGCGGTGGCGCCACCAATGGCAGACATAACGAAGCGCCGCGAGTGGGT 5286

Qy 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThr 40
Db AATGCTTCAGGAAATTTGGCATTTCCCATGGCTGGCGCAGACAGAGTCATCACCACC 5346

Qy 41 SerThrArgThrTrpAlaLeuProThrTrpAenAenHisLeuTyrlsGlnIleSerSer 60
Db AGCACCCGACCTGGGCGCTTGGCCACCTACAAATACCACTCTACAGCAAAATCTCCAGT 5406

Qy 61 AlaSerThrGlyAlaSerAsnAspAenHisTyrlsPheGlyTyrlsSerThrProTrpGlyTyrls 80
Db GCTTCACAGGCGGCCGCGACGACCAACCACTACTTCGGCTACAGCACCCCTGGGGGTAT 5466

Qy 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAen 100
Db TTTGATTTCAACAGATTCACCTGGCCACTTTTCCACCACTGGTGGCAGCGACTCATCAAC 5526

Qy 101 AsnAenTrpGlyPheArgProIysArgLeuAenPheLysLeuPheAenIleGlnValLys 120
Db AACAAATGGGAGATTCGGGCCCAAGAGACTCAACTTCAAACTCTTCAACATCAAGTCAAG 5586

Qy 121 GluValThrThrAsnAspGlyValThrThrIleAlaAenAenLeuThrSerThrValGln 140
Db GAGGTACAGACGAATGATGGGTCAACCACTCGCTATATACCTTACCAAGCAGCGTTCA 5646

Qy 141 ValPheSerAspSerGlyTrpGlnLeuProTyrlsValLeuGlySerAlaHisGlnGlyCys 160
Db GTCTTCTCGGACTCGGAGTACAGCTTCGGTAGCTCTCGGCTCGCGCAGCAGCGGCTGC 5706

Qy 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrlsGlyTyrlsLeuThrIleAen 180
Db CTCCCTCGCTTCCCGCGCGAGAGTGTTCATGATTCGCAATACGGCTACTCAGCTCAAC 5766

Qy 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrlsCysLeuGluTyrlsPheProSerGln 200
Db AATGGCAGCCCAAGCGTGGGAGCTTCATCTTTTACTGCTGGATATTTCCCTTCTCAG 5826

Qy 201 MetLeuArgThrGlyAsnAspPheThrPheSerTyrlsThrPheGluGluValProPheHis 220
Db ATGCTGAGAACGGGCAACAACCTTTACCTTCAGTACACCTTTGAGGAAGTGCCTTTCCAC 5886

Qy 221 SerSerTyrlsAlaHisSerGlnSerLeuAspArgLeuMetAenProLeuIleAspGlnTyrls 240
Db AGCAGCTACGGCAGCAGCAGCGCTGACCGGCTGATGAAATCTCTCATGACCAATAC 5946

Qy 241 LeuTyrlsTyrlsLeuAsnArgThrGlnAenGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
Db CTGTATTTACCTGAACAGAACTCAAAATCAGTCCGGAAGTGGCCAAACAGAGACTTGTG 6006

Qy 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
Db TTTAGCCGTGGGTCTCCAGCTGGCATGCTGTTCAGCCCAAAACCTGGCTACCTGGAGCC 6066

Qy 281 CysTyrlsArgGlnArgValSerIysThrIysThrAspAenAenAenSerAsnPheThr 300
Db TGTATTCGCGACGCGGCTTTCTAAAACAAAAACAGACAAACAGCAATTTTACC 6126

Qy 301 TrpThrGlyAlaSerLysTyrlsAenLeuAenGlyArgGluSerIleIleAenProGlyThr 320
Db TGGACTGGTCTTCAAAATATTAACCTCAATGGGCGTGAATCCATCATCAACCTGGCACT 6186

Qy 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
Db GCTATGGCTTCACAAAGACGACGAAAGCAAGATTTCTTCCATGAGCGGTGTCATGATT 6246

Qy 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAenValMetIleThrAsp 360
Db TTTGAAAGAGAGCGCGGAGCTTCAACACATGTCATGTCATGATGATTTACAGAC 6306

Qy 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
Db GAAGAGGAATTAAGACCACTTACCCCTGTGCCACCCGAGAGATTTGGGACCGTGGCAGTC 6366

Qy 381 AsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
Db AATTTCCAGAGCAGCAGCAGACCCCTGCGACCCGAGATGTGCATGTCTATGGGAGCATTA 6426

Qy 401 ProGlyMetValTrpGlnAspArgAspValTyrlsLeuGlnGlyProIleTrpAlaValIle 420
Db CCTGGCATGGTGTGGCAAGATAGAGACGTGTACCTGAGGGTCCCATTTGGGCCCAAAATT 6486

Qy 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAen 440
Db CCTCACAGATGGACACTTTCACCCGCTCTCTTATGGCGGCTTTGGACTCAAGAAC 6546

Qy 441 ProProGlnIleLeuLysAenThrProValProAlaAenProAlaGluPhe 460
Db CCGCTCTCAGATCTCTCATCAAAACACGCGCTTCTCTGCGAATCTCTCCGCGGAGTTT 6606

Qy 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrlsSerThrGlyGlnValSerValGlu 480
Db TCAGCTACAAAGTTTGTCTTCAATCACCCTTACTCCACGAGCAAGTGGTGTGGNA 6666

Qy 481 IleGluTrpGluLeuGlnLysGluAenSerLysArgTrpAenProGluValGlnTyrlsThr 500
Db ATGAATGGGAGCTGTCAGAAAGAAACAGCAGCGCTGGAATCCCGAAGTGCAGTACACA 6726

Qy 501 SerAsnTyrlsAlaLysSerAlaAenValAspPheThrValAspAenAsnGlyLeuTyrlsThr 520
Db TCCAAATATGCAAAATCTGCCACGTTGATTTTACTGTGGACAAACAATGGGACTTTTACT 6786

Qy 521 GluProArgProIleGlyThrArgTyrlsLeuThrArgProLeu 534
Db GAGCTCGCCCCCATTTGGCACCCGCTTACTTACCCGTCCTCCCTG 6828

RESULT 3
US-10-291-583-26
; Sequence 26, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alvira, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; TITLE OF INVENTION: Sequences and Isolating No. US20030138772A1
; FILE REFERENCE: US-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; PRIOR FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,086
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26
; LENGTH: 3075
; TYPE: DNA
; ORGANISM: new AAV serotype, clone H2
US-10-291-583-26

Alignment Scores:

```
Pred. No.: 1,49e-285 Length: 3075
Score: 2536.50 Matches: 455
Percent Similarity: 91.57% Conservative: 34
Best Local Similarity: 85.21% Mismatches: 44
Query Match: 87.28% Indels: 1
DB: 13 Gaps: 1

US-09-807-802A-17 (1-534) x US-10-291-583-26 (1-3075)

QY 1 MetAlaSerGlyGlyGlyValaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGly 20
DB 1409 ATGGCTACAGCGCAGTGGCGCAACCAATGGCAGACAATAACGAGGCTGCGAGTGGGT 1468
QY 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThr 40
DB 1469 AATTCCTCAGGAATTTGGCATTTGCCAATGGCTGGCGCAGAGTCAATCACCACC 1528
QY 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSer 60
DB 1529 AGCACCCGNACTGGGCCCTGCCACATACACACACCCTCTACAGCAATCTCCAGC 1588
QY 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
DB 1589 CAATCA---GGAGCCAGCAACGACCACTACTTTGGCTACAGCACCCCTGGGGGTAT 1645
QY 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
DB 1646 TTTCAGCTTCAACAGATTCCACTGCCACTTTTCCACACGCTGACTGGCAAGACTCATCAAC 1705
QY 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 120
DB 1706 AACAACTGGGGATTCCGGCCCAAAAGACTCACTTCAAGCTCTTAAATATCAAGTCAA 1765
QY 121 GluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeuThrSerThrValGln 140
DB 1766 GAGGTCAACAGAAATGACGGTACGACGAGATTGGCAATTAACCTTACCAGACGGTTTCA 1825
QY 141 ValPheSerAspSerGlnTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
DB 1826 GTGTTTACTGACTCGGAGTACAGACTCCCGTACGCTGGGCTGGCGCATCAAGATGC 1885
QY 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
DB 1886 CTCCTCCGCTTCCAGCGAGCTTTCATGGTCCACAGTATGGATACCTCAACCTGAAAC 1945
QY 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
DB 1946 AACGGGAGTCAGCGGTAGGACGCTCTTCTTTTACTGCTGGAGTACTTTCTCTTCTCAG 2005
QY 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
DB 2006 ATGCTCGGTACTGGAAACAACTTTCAGTTGAGTACACATTTTGAAGACGTGCTTTCCAC 2065
QY 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
DB 2066 AGCAGCTACGCTCACAGCCAGAGTCTGGATCGGCTGATGAATCTCTGTATGACCACTAC 2125
QY 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
DB 2126 CTGTATTATCTGAACAGACACAAACAAATAGTGGAACTCTTCCAGAGTCTCGGCTACTG 2185
QY 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
DB 2186 TTTGAGCAAGCTGGCAACCAACATGTCTTCAAGCTTAAAACTGGCTGGCTGGACCT 2245
QY 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
DB 2246 TGCTACAGACAGCAGCGTCTGTCAAACAGGGCAACGACAAACAAACAGCAACTTTCC 2305
QY 301 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr 320
DB 2306 TGGAGCTGCAGCTACAAAGATATCATCTAAATGGCGGGGACTCGTGTGTTTAAATCAGGACCA 2365
```

```
QY 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
DB 2366 GCTATGGCCAGTCAACAGGATGACGAAGAAAGTTTTTCCCATGTCATGGACCCCTGATA 2425
QY 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
DB 2426 TTTGGTAAACAAGGAACAAATGCCACGCGGATTTGGAAAATGTTCATGATTACAGAT 2485
QY 361 GluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
DB 2486 GAAGAGAATAATCAGGGCCACCAATCCGCTGGCTACGAGCAGTAGTCGGACACTGTGTCAAAT 2545
QY 381 AsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
DB 2546 AATTTGCCAATACTCAACACTGGTCCAACTACTGGAACGTCTCAATCGCAAGGAGCGTTA 2605
QY 401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 420
DB 2606 CCTGGTATGGTGTGGCAGGATCGAGACGTGTACTCCTGAGGAGCCCATTTGGGCCAAGATT 2665
QY 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
DB 2666 CCTCACCCGATGACACTTTTCATCCTTCTCCACTGATGGAGGTTTTTGGACTCAAAACRC 2725
QY 441 ProProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460
DB 2726 CCGCTCTCTAGATCATGATCAAAACACTCCCGTCCAGGCCAATCTCTCCCAAACTTC 2785
QY 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
DB 2786 AGTTTGTCCAAAGTTTGTCTTCTTCATCACAGATTTTCCACGGGACAGGTGAGCGTGGAG 2845
QY 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrThr 500
DB 2846 ATCAGTGGGAGCTGCGAGAGGAGAACAGCAACGCTGGAATCCGGAATTCGGAATTCAGTACT 2905
QY 501 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 520
DB 2906 TCCAACTACAACAAGTCTGTTAATGTGGACTTTACTGTGGACACTAATGTTGTGTATTCA 2965
QY 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
DB 2966 GAGCCTCGCCCATTTGGCACCAGATACCTGACTCTGTAATCTG 3007
```

RESULT 4

```
US-10-291-583-8
; Sequence 8, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alvira, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; FILE REFERENCE: UPN-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 4726
; TYPE: DNA
; ORGANISM: adeno-associated virus serotype 3
US-10-291-583-8
```

Alignment Scores:
Pred. No.: 2, 93e-284 Length: 4726
Score: 2528.00 Matches: 459
Percent Similarity: 91.03% Conservative: 28
Best Local Similarity: 85.79% Mismatches: 46
Query Match: 86.99% Indels: 2
DB: 13 Gaps: 2

US-09-807-802A-17 (1-534) x US-10-291-583-8 (1-4726)

QY 1 MetAlaSerGlyGlyAlaProMetAlaAsePheAseGluGlyAlaAseGlyValGly 20
DB 2815 ATGGCTTCAGGCGGTGGCGCCCAATGCGACAAATACAGAGGTGGCGATGGGT 2874
QY 21 AenAlaSerGlyAseThrPheCysAsePheThrTrpLeuGlyAsePheValThr 40
DB 2875 AATTCCTCAGGAAATGGCAATGGCAATGCCATGGCTGGCGGACAGAGTCATCACACC 2934
QY 41 SerThrArgThrTrpAlaLeuProThrTrpAseAseHisLeuTyrLysGlnAseSer 60
DB 2935 AGCACCAGAACCTGGGCCCTGCCACTTACAAACCACTCTCTCAAGCAAAATCTCCAGC 2994
QY 61 AlaSerThrGlyAlaSerAseAseHisThrPheGlyTyrSerThrProThrGlyTyr 80
DB 2995 CAATCA---GGAGCTTCAACGACAAACCACTACTTTGGCTACAGCACCCCTGGGGGTAT 3051
QY 81 PheAsePheAseAseThrPheHisCysAsePheProArgAseThrGlnArgLeuLeuAse 100
DB 3052 TTTGACTTTAACAGATTCCATGCCACTTCTCACCAGCTGATGGCGGACGACTCATTAAC 3111
QY 101 AseAseThrGlyPheArgProLysArgLeuAsePheLysLeuPheAseGlnValLys 120
DB 3112 AACAACTGGGATTCGGGCCCAAGAACTCAGCTTCAAGCTCTTCAACATCAAGTTAGA 3171
QY 121 GluValThrAseAseGlyValThrThrLeuAseAseLeuThrSerThrValGln 140
DB 3172 GGGGTTCACGACAGATGGCGACGACTATTCGCAATAACCTTACCAGCACGGTTCA 3231
QY 141 ValPheAseAseSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
DB 3232 GTGTTTACGGATCGGATATCAGCTCCGTCAGTCTCGGTCGGCGCACCAAGGCTGT 3291
QY 161 LeuProProPheProAlaAsePheMetIleProGlnTyrGlyTyrLeuThrLeuAse 180
DB 3292 CTCCCGCGCTTCCAGCGGACGCTTTCATGGTCCCTCAGTATGGATACCTCACCCCTGA 3351
QY 181 AseGlySerGlnAlaValGlyArgSerPheTyrCysLeuGluTyrPheProSerGln 200
DB 3352 AACGGAAGTCAACGCGTGGCGGCTCATCTTTTACTGCTGGAGTACTTCCCTTCGCGAG 3411
QY 201 MetLeuArgThrGlyAseAsePheThrPheSerTyrThrPheGluGluValProPheHis 220
DB 3412 ATGCTTAAGGACTGGAAATAACTTCCAACTACGTATACCTTCGAGATGTACTTTTAC 3471
QY 221 SerSerTyrAlaHisSerGlnSerLeuAsePheArgLeuMetAseProLeuAsePheGlnTyr 240
DB 3472 AGCAGCTACGCTCACAGCCAGAGTTGGATCGCTTGGATGAACTCTTATTGATCAGTAT 3531
QY 241 LeuTyrTyrLeuAseArgThrGln---AseGlnSerGlySerAlaGlnAseLysAseLeu 259
DB 3532 CTGTACTACTCAACGAGAACGCAAGAAACAACCTCTGGAACCAACAACCAATCACGGCTG 3591
QY 260 LeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAseThrLeuProGly 279
DB 3592 CTTTATGACCGCTGGCGCTCAGTCTATGCTTTTGCAGGCCAGAAATTTGGTACTGGG 3651
QY 280 ProCysTyrArgGlnGlnArgValSerTyrThrLysThrAseAseAseAseSerAsePhe 299
DB 3652 CCCTGCTACCGGCAACAGAGACTTTCAAGAGCTGTAAACGACAAACCAACAGTAATTT 3711
QY 300 ThrTrpThrGlyAlaSerLysTyrAseLeuAseGlnArgGluSerIleAseProGly 319

DB 3712 CCTTGACAGCGCCAGCAAAATATCATCTCAATGGCGGAGCTGCTGGTGAATCCAGGA 3771
QY 320 ThrAlaMetAlaSerHisLysAsePheGluAsePhePhePheProMetSerGlyValMet 339
DB 3772 CAGCTATGGCCAGTTCACAGGACGATGAAGAAATTTTCTTATGACGCGCAATCTA 3831
QY 340 IlePheGlyLysGluSerAlaGlyAlaSerAseThrAlaLeuAsePheAseValMetIleThr 359
DB 3832 ATATTTGGCAAGAGGACACAGCGCAAGTAACGCAAGATTAGATAATGATTAATGATACG 3891
QY 360 AspGluGluLeuLysAlaThrAseProValAlaThrGluArgPheGlyThrValAla 379
DB 3892 GATGAAGAGAGATTGCTACCAATCTCTGTGGCAACAGACGATGATGAACCTGTGGCA 3951
QY 380 ValAsePheGlnSerSerSerThrAseProAlaThrGlyAsePheValHisAlaMetGlyAla 399
DB 3952 AATAACTTGCAGAGCTCAATACAGCTCCACGACTGGAATGTCAATCATCATCAGGGGCC 4011
QY 400 LeuProGlyMetValTrpGlnAsePheArgValTyrLeuGlnGlyProIleThrAlaLys 419
DB 4012 TTACTTGGCATGGTGTGGCAAGATCGTGACGTGTACCTTCAAGGACCTATCTGGGCAAG 4071
QY 420 IleProHisThrAsePheGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLys 439
DB 4072 ATTCTCACAGGATGGACACTTTTCATCTCTCTCTGATGGAGGCTTTGGACTGAAA 4131
QY 440 AseProProGlnIleLeuLysAseThrProValProAlaAseProProAlaGlu 459
DB 4132 CATCCGCTCTCAATCATGATCAAAATACTCCGTCACGCAATCTCTCCGACGACT 4191
QY 460 PheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerVal 479
DB 4192 TTCAGCCGCGCAAGTTGCTTTCATTTATCACTCAGTACTCCACTGGACAGGTGAGGCTG 4251
QY 480 GluIleGluTrpGluLeuGlnLysGluAseSerLysArgTrpAseProGluValGlnTyr 499
DB 4252 GAAATTTGAGTGGGAGCTACAGAAAGAAACAGCAACCTTGGAAATCCAGAGATTCAGTAC 4311
QY 500 ThrSerAseTyrAlaLysSerAlaAseValAsePheThrValAseAseAseGlyLeuTyr 519
DB 4312 ACTTCAACTACAAAGTCTGTTAATGTGAGCTTTACTGTAGACACTTAATGGTGTAT 4371
QY 520 ThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
DB 4372 AGTGAACCTCGCCCTATTGGAACCCGCTATCTCACAGAAACTTG 4416

RESULT 5

US-10-291-583-25/c
; Sequence 25, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alvirra, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; FILE REFERENCE: US 60/341,117
; FILE REFERENCE: US 60/341,117
; CURRENT APPLICATION NUMBER: US/10/291,583
; PRIOR FILING DATE: 2002-11-12
; PRIOR FILING DATE: 2002-11-12
; PRIOR FILING DATE: 2001-11-13
; PRIOR FILING DATE: 2001-12-17
; PRIOR FILING DATE: 2001-12-17
; PRIOR FILING DATE: 2002-05-01
; PRIOR FILING DATE: 2002-05-01
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 25
; LENGTH: 3142
; TYPE: DNA
; ORGANISM: new AAV serotype, clone H6

```

US-10-291-583-25
Alignment Scores:
Pred. No.: 1,15e-283 Length: 3142
Score: 2520.50 Matches: 453
Percent Similarity: 91.39% Conservative: 35
Best Local Similarity: 84.83% Mismatches: 45
Query Match: 86.73% Indels: 1
DB: 13 Gaps: 1

US-09-807-802A-17 (1-534) x US-10-291-583-25 (1-3142)
Qy 1 MetAlaSerGlyGlyAlaProMetAlaAspAenAngluGlyAlaAspGlyValGly 20
Db 1733 ATGGCTACAGGCAGTGGCGCCACCAATGGCAGACAATAACGAGGGTGCAGTGGAGTGGGT 1674
Qy 21 AenAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThr 40
Db 1673 AATTCCTCAGGAATTTGGCATTTGGATTTCCCAATGGCTGGCGCAGAGTCATCACCACC 1614
Qy 41 SerThrArgThrTrpAlaLeuProThrTyrAenAenHisLeuTyrLysGlnIleSerSer 60
Db 1613 AGCACCAGCACTGGGCCCTGCCCAATACACCAACCACTCTACAGCAAAATCTCCAGC 1554
Qy 61 AlaSerThrGlyAlaSerAenAenHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
Db 1553 CAATCA---GGAGCCAGCAGCAACCACTACTTTGGCTCAGCACCCTCTGGGGGTAT 1497
Qy 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAen 100
Db 1496 TTTGACTTCAACAGATTCCATGCCCTTTTCCACCACTGACTGGCAAGACTCGTCAAC 1437
Qy 101 AenAenTrpGlyPheArgProLysArgLeuAenPheLysLeuPheAenIleGlnValLys 120
Db 1436 AACAAACCGGGGATTCGGGCCCAAAAGACTCAACTTCAAGCTCTTAATATTCAGTCAA 1377
Qy 121 GluValThrAenAspGlyValThrThrIleAlaAenAenLeuThrSerThrValGln 140
Db 1376 GAGGTCAACGAGATGACGGTACGAGCAGATTTGCCATTAACCTTACCGACAGGTTGAG 1317
Qy 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
Db 1316 GTGTTTACTGACTCGGAGTACAGCTCCCGTAGCTCTGGCTCGCGCATCAAGGATGC 1257
Qy 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAen 180
Db 1256 CTCCCGCCGTTTCCAGCGCAGCTTTCATGGTCCCAAGATATGGATACCTCACCCCTGAAC 1197
Qy 181 AenGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
Db 1196 AACGGAGTCAAGCGGTAGACGCTCTTCTTTTACTGCTGTGAGTACTTCTTCTCTGAG 1137
Qy 201 MetLeuArgThrGlyAenAenPheThrPheSerTyrThrPheGluGluValProPheHis 220
Db 1136 ATGCTCGGTACTTGGAAACCACTTTTCAAGTTCAGTACACTTTTGAAGACGCTGCTTCCAC 1077
Qy 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAenProLeuIleAspGlnTyr 240
Db 1076 AGCAGTACGCTCACAGCCAGAGTCTGGATCGGCTGATGAATCTCTGTATCGACCACTAC 1017
Qy 241 LeuTyrTyrLeuAenArgThrGlnAenGlnSerGlySerAlaGlnAenLysAspLeuLeu 260
Db 1016 CTGTATTATCTGAACAGACACACAAATAAGTAGTGGAACTCTTCAGCAGTCTCGGTACTG 957
Qy 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAenTrpLeuProGlyPro 280
Db 956 TTTAGCAAGCTGGACCAACCAATGTCTCTTCAAGCTAAATACTGGCTGCTGCTGACCT 897
Qy 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAenAenSerAenPheThr 300
Db 896 TGCTACAGACAGCAGCGCTCTGTCAAAACAGCGCAACACCAACCAACCACTTCC 837
Qy 301 TrpThrGlyAlaSerLysTyrAenLeuAenGlyArgGluSerIleIleAenProGlyThr 320

```

```

Db 836 TGGACTCCAGCTACAAAGTATCATCTAATGGCGGAGCTCGTTGGTTAATCCAGGACCA 777
Qy 321 AlaMetAlaSerHisLysAspAspGluAspPhePhePheProMetSerGlyValMetIle 340
Db 776 GCTATGGCCAGTCACAGGATGACGAAGAAAGTTTTCCTCCATGTCATGGAACCTTGATA 717
Qy 341 PheGlyLysGluSerAlaGlyAlaSerAenThrAlaLeuAenAenValMetIleThrAsp 360
Db 716 TTTGGTAAACAGGAACAATGCCAGCGGATTGGAAATGTCATGATTACAGAT 657
Qy 361 GluGluGluIleLysAlaThrAenProValAlaThrGluArgPheGlyThrValAlaVal 380
Db 656 GAAGAAGAATCAGGCGCCACCAATCCCGTGGCTACGAGCAGTACGGGACTGTGTCAAT 597
Qy 381 AenPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
Db 596 AATTTGCAAAACTCAAACTACTGGTCCAACTACTGGAACCTGTCAATCACCAGGAGCGTTA 537
Qy 401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 420
Db 536 CCTGGTATGGTGTGGCAGGATCGAGAGCTGTACCTCGAGGAGCCCATTTGGGCCCAAGAT 477
Qy 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAen 440
Db 476 CCTCAGCCAGTGGACACTTTCATCTTCTCCACTGATGGAGGTTTGGACTCAAAACAC 417
Qy 441 ProProGlnIleLeuLysAenThrProValProAlaAenProAlaGluPhe 460
Db 416 CCGCTCTCAGATCATGATCAAAACACTCCCGTTCAGCCCAATCTCTCCCAAACTTC 357
Qy 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
Db 356 AGTTCTGCCAAGTTTGTCTTTTATCACAGTATTTCCAGCGACAGTTCAGCGTGGAG 297
Qy 481 IleGluTrpGluLeuGlnLysGluAenSerLysArgTrpAenProGluValGlnTyrThr 500
Db 296 ATCCAGTGGGAGCTGCAGAGGAGAACAGCAACCGCTGGAATCCGAAATTCAGTACACT 237
Qy 501 SerAenTyrAlaLysSerAlaAenValAspPheThrValAenAenGlyLeuTyrThr 520
Db 236 TCCAACTACAAAGTCTGTTAATGTGGACTTTTACTGTGGACACTAATGGTGTGTATTCA 177
Qy 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
Db 176 GAGCTCGCCCATTTGGCACCAGATACCTGACTCGTAACTCTG 135

RESULT 6
US-10-205-942-5
; Sequence 5, Application US/10205942
; Publication No. US20030053990A1
; GENERAL INFORMATION:
; APPLICANT: University of No. US20030053990A1th Carolina-Chapel Hill
; APPLICANT: Rabinowitz, Joseph E.
; APPLICANT: Samulski, Richard J.
; APPLICANT: Xiao, Weidong
; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING THE SAME
; FILE REFERENCE: 5470-186
; CURRENT APPLICATION NUMBER: US/10/205,942
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 8179
; TYPE: DNA
; ORGANISM: Adeno-associated virus
US-10-205-942-5
Alignment Scores:
Pred. No.: 1,04e-282 Length: 8179
Score: 2518.00 Matches: 458
Percent Similarity: 90.65% Conservative: 27
Best Local Similarity: 85.61% Mismatches: 48

```

Query Match: 86.65% Indels: 2
DB: 15 Gaps: 2
US-09-807-802A-17 (1-534) x US-10-205-942-5 (1-8179)

Qy 1 MetAlaSerGlyGlyGlyAlaProMetAlaAspAenAenGluGlyAlaAspGlyValGly 20
Db ATGGCTACAGCGAGTGGCGCCACCAATGCGACAAATAACGAGGGCGCGACGAGTGGGT 2798

Qy 21 AenAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThr 40
Db AATTCCTCGGNAATGGCATTTGCCATTCACATGGATGGCGACAGAGTCAATCAACC 2858

Qy 41 SerThrArgThrTrpAlaLeuProThrTyrAenAenHisLeuTyrLysGlnIleSerSer 60
Db AGCACCAGAACCTGGCGCCCTGCCACCTACAAACACCACTCTCAAAACAAATTCACAG 2918

Qy 61 AlaSerThrGlyAlaSerAenAspAenHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
Db CAATCA---GGAGCCTCGAACACCAATCACTACTTTTGGCTACAGCACCCCTTGGGGGTAT 2975

Qy 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAen 100
Db TTTGACTTCACAGATTCCACTGCCACTTTTACCACCGTGAATGGCAAGACTCATCAAC 3035

Qy 101 AenAenTrpGlyPheArgProLysArgLeuAenPheLysLeuPheAenIleGlnValLys 120
Db AACAACTGGGGATTCCGACCCCAAGAGACTCAACTTCAAGCTCTTTAAACATTTCAAGTCAA 3095

Qy 121 GluValThrThrAenAspGlyValThrThrIleAlaAenAenLeuThrSerThrValGln 140
Db GAGGTCACGCGAATGACGGTACGACGACGATGTCCTAATAACCTTACCAGCACCGTTACG 3155

Qy 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
Db GTGTTTACTGACTCGAGTACCACTCCGCTAGCTCGGCTCGGCGGACCAAGGCTGT 3215

Qy 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAen 180
Db CTCCGCGGCTTTCCACGCGAGCTCTTCATGGTCCCTCAGTATGATGATCACTCACCTGAAC 3275

Qy 181 AenGlySerGlnAlaValGlyArgSerPheTyrCysLeuGluTyrPheProSerGln 200
Db AACGAAAGTCAAGCGGTGGAGCGCTCATCTTTTACTGCTGGAGTACTTCCCTTCGCGAG 3335

Qy 201 MetLeuArgThrGlyAsnAenPheThrPheSerTyrThrPheGluGluValProPheHis 220
Db ATGCTAAGGACTGGAAATAACTTCCAATTCACTATACCTTCGAGGATGATACCTTTTAC 3395

Qy 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAenProLeuIleAspGlnTyr 240
Db AGCAGTACGCTCACAGCCAGAGTTGGATCGCTTGAATGATCACTCTTATTTGATCAGTAT 3455

Qy 241 LeuTyrTyrLeuAenArgThrGln---AenGlnSerGlySerAlaGlnAenLysAspLeu 259
Db CTGTACTACCTGAAACAGAACGCAAGAACCACTCTGGAACACCAACCAATACCGGCTG 3515

Qy 260 LeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAenTrpLeuProGly 279
Db CTTTITAGCCAGGCTGGGCTCAGCTATGCTTTTGGAGGCGAGAAATGGCTACCTGGG 3575

Qy 280 ProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAenAenAenSerAenPhe 299
Db CCCTCTACCGGCAACAGAGACTTTCAAAGACTGCTAAACGACCAACCAACAGCAATCACTT 3635

Qy 300 ThrTrpThrGlyAlaSerLysTyrAenLeuAenGlyArgGluSerIleIleAenProGly 319
Db CCTTGGACAGCGGCCAGCAAAATATATCTCAATGGCCGCGACTCGCTGGTAATCCAGGA 3695

Qy 320 ThrAlaMetAlaSerHisLysAspAspGluAenLysPhePheProMetSerGlyValMet 339
Db CCAGCTATGGCCAGTCAACAGGACCATGAAGAAAAATTTTCCCTATTCACGCGCAATCTA 3755

Qy 340 IlePheGlyLysGluSerAlaGlyAlaSerAenThrAlaLeuAenAenValMetIleThr 359
Db ATATTTGCAAGAAAGGCAACACGCAAGTAACGACAGATTTAGATAATGTAAATGATTACG 3815

Qy 360 AspGluGluLeuLysAlaThrAenProValAlaThrGluArgPheGlyThrValAla 379
Db GATGAAGAAGAGATTCTGACCACCAATCTGTGGCAACAGACAGTATGGAACTGTGGCA 3875

Qy 380 ValAenPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAla 399
Db AATAACTTGCAGAGCTCAATAACAGCTCCACGACTGGAACCTGTCAATCATCAGCGGGCC 3935

Qy 400 LeuProGlyMetValTrpGlnAenAspArgAspValTyrLeuGlnGlyProIleTrpAlaLys 419
Db TTACTCTGGCATGGTGTGGCAAGATCGTACGTGTACCTTCAAGGACCTATCTGGGCAAG 3995

Qy 420 IleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLys 439
Db ATTCTCTCACGAGATGGACACTTTTCATCTCTCTCTGATGGAGGCTTTGGACTGAAA 4055

Qy 440 AsnProProGlnIleLeuIleLysAenThrProValProAlaAenProProAlaGlu 459
Db CATCGGCTCTCTCAATCATGATCAAAAATACCTCGGTACTCGGAATCTCTTCACCAACC 4115

Qy 460 PheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerVal 479
Db TTCAGTGCGCAAAAGTTTGTCTTCTTCATCACACAGTACTCCACGGACAGGTGAGGTG 4175

Qy 480 GluIleGluTrpGluLeuGlnLysGluAenSerLysArgTrpAenProGluValGlnTyr 499
Db GAGATCAGATGGGAGCTGCGAGAGAAACAGAAACGCTGGAAATCCGAAATTCAGTAC 4235

Qy 500 ThrSerAsnTyrAlaLysSerAlaAenValAspPheThrValAspAenAenGlyLeuTyr 519
Db ACTTCCAACTAGAACAAAGTCTGTTAATCGTGGACTTACCGTGGATACTAATGGCGGTAT 4295

Qy 520 ThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
Db TCAGAGCTCGCCCCATTGGCACACAGATCTGACTGACTGTAATCTG 4340

RESULT 7
US-09-804-898-1
; Sequence 1, Application US/09804898
; Patent NO. US20020045264A1
; GENERAL INFORMATION:
; APPLICANT: XIAO, WEIDONG
; TITLE OF INVENTION: PRODUCTION OF CHIMERIC CAPSID VECTORS
; FILE REFERENCE: 102182-14
; CURRENT APPLICATION NUMBER: US/09/804,898
; CURRENT FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 60/189,110
; PRIOR FILING DATE: 2000-03-14
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; TYPE: DNA
; ORGANISM: adeno-associated virus 2
US-09-804-898-1

Alignment Scores:
Pred. No.: 7,92e-279 Length: 4679
Score: 2481.50 Matches: 445
Percent Similarity: 90.26% Conservative: 37
Best Local Similarity: 83.33% Mismatches: 51
Query Match: 85.39% Indels: 1
DB: 9 Gaps: 1

US-09-807-802A-17 (1-534) x US-09-804-898-1 (1-4679)

Qy 1 MetAlaSerGlyGlyGlyAlaProMetAlaAspAenAenGluGlyAlaAspGlyValGly 20
Db ATGGCTACAGCGAGTGGCGCCACCAATGCGACAAATAACGAGGGCGCGACGAGTGGGT

Db 2809 ATGGCTACAGGCGAGTGGCGCCACCAATGGCAGACAAATACAGGGCGCCGACGAGTGGGT 2868
Qy 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 40
Db 2869 AATTCCTCGGAAATGGCATTTGGCATTCACATGGATGGCGCAGAGTCAATCACCACC 2928
Qy 41 SerThrArgThrTrpAlaLeuProThrTrpAsnAsnHisLeuTyrGlnIleSerSer 60
Db 2929 AGACCCCGAACTGGGCGCTGCCACCTTACAAACACCACTCTTACAAACAAATTTCCAGC 2988
Qy 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
Db 2989 CAATCA--GGAGCTCGAACGACCAATCACTACTTTGGCTACAGCACCCCTTGGGGGTAT 3045
Qy 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
Db 3046 TTTGACTTCAACAGATTCCACTGCCACTTTTACCACGCTGACTGGCAAGACTCATCAAC 3105
Qy 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLeuPheAsnIleGlnValLys 120
Db 3106 AACAACTGGGGATTCCGACCCCAAGAGACTCAACTTCAAGCTCTTTAAACATTCAAGTCAAA 3165
Qy 121 GluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeuThrSerThrValGln 140
Db 3166 GAGGTCAACGAGAAATCAAGTACGAGCAGGATTTGCCAATACTTACCAGCAGGTTCAG 3225
Qy 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
Db 3226 GTGTTTACTGACTCGGAGTACCAGCTCCGCTAGCTCTCGGCTCGGCGCATCAAGGATGC 3285
Qy 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
Db 3286 CTCGCCCGCTTCCAGCAGAGCGTCTTCATGGTGCCACAGTATGGATACCTCACTCCCTGAAC 3345
Qy 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
Db 3346 AACGGGAGTCAGGAGTAGGACGCTCTTCAATTTACTGCTGGAGTACTTCTCTTCTCAG 3405
Qy 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
Db 3406 ATGCTCGTACCGGAAACAACTTTTACCTTCAGCTACACTTTTGGAGACGCTCTCTTCCAC 3465
Qy 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
Db 3466 AGCAGCTAGCTCAGCAGCAGAGTCTGGACGCTCTCATGAATCTCTCATCAGCAGTAC 3525
Qy 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
Db 3526 CTGTATTACTTGAGCAGAACAAACACTCCCAAGTGGAAACCAACCAACGAGTCAAGGCTTCAG 3585
Qy 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
Db 3586 TTTTCTCAGCGCCGAGCGAGTGACATTCGGGACCACTAGGAACTGGCTTCTCTGACCC 3645
Qy 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsnPheThr 300
Db 3646 TGTATTCCGCGCAGCGAGTATCAAGACATCTCGCGGATACCAACACAGTGAATCTCG 3705
Qy 301 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr 320
Db 3706 TGGACTGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTGTGTAATCGGGCCCG 3765
Qy 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
Db 3766 GCCATGGCAAGCCACCAAGGACGATGAAGAAAGTTTTTCTCAGAGCGGGGTCTCATC 3825
Qy 341 PheGlyLysGlySerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
Db 3826 TTTGGGAACCAAGGCTCAGAGAAACAAATGTGGACATGTAAGAGTCAATGATCAGAC 3885
Qy 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
Db 3886 GAAGAGAAATCAGGCAACCAATCCCGTGGCTACGAGCAGTATGGTTCTGTATCTACC 3945

Qy 381 AsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
Db 3946 AACCTCCAGAGAGCAACAGACAGCTACCCAGATGTCAACACACAAAGGCGTTCTT 4005
Qy 401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 420
Db 4006 CCAGGCATGGTCTGGCAGGACAGAGATGTGTACCTTCAGGGGCGCCATCTGGGCAAGATT 4065
Qy 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
Db 4066 CCACACAGGACGAGCATTTTACCCCTCTCCCTCATGGTGGATTCGGACTTTAAACAC 4125
Qy 441 ProProProGlnIleLeuIleLysAsnThrProValProAlaAsnProAlaGluPhe 460
Db 4126 CCTCTCCACAGATTCATCAAGAACACCCCGGCTACTCGGAATCTCTCCACACCTTC 4185
Qy 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
Db 4186 AGTGGGCAAGTTTGTCTTCTCATCACAGTACTCCACGGCAGAGTCAGCGTGGAG 4245
Qy 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrThr 500
Db 4246 ATCAGTGGGAGCTGCAGAGGAAACAGCAACGCTGGAAATCCGAAATTCAGTACACT 4305
Qy 501 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 520
Db 4306 TCCAACTACAAACAGTCTGTAAATGTGACTTTTACTGTGGACATAATGGCGTGTATTCA 4365
Qy 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
Db 4366 GAGCTCGCCCAATTGGCACCAGATACCTGACTCGTAATCTG 4407

RESULT 8

US-09-945-681-10
; Sequence 10, Application US/09945681
; Patent No. US20020064878A1
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITE DE NANTES
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR RECOMBINANT
; FILE REFERENCE: B4182AA - UNIVERSITE DE NANTES
; CURRENT APPLICATION NUMBER: US/09/945,681
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: PCT/EP 00/01854
; PRIOR FILING DATE: 2000-03-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 4679
; TYPE: DNA
; ORGANISM: adeno-associated virus 2
US-09-945-681-10

Alignment Scores:
Pred. No.: 7,92e-279 Length: 4679
Score: 2481.50 Matches: 445
Percent Similarity: 90.36% Conservative: 37
Best Local Similarity: 83.33% Mismatches: 51
Query Match: 85.39% Indels: 1
DB: 9 Gaps: 1

US-09-807-802A-17 (1-534) x US-09-945-681-10 (1-4679)

Qy 1 MetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGlyAlaAspGlyValGly 20
Db 2809 ATGGCTACAGCAGTGGCGCCCAATGGCAGACAAATACAGGGCGCCGACGAGTGGGT 2868
Qy 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 40
Db 2869 AATTCCTCGGAAATGGCATTTGGCATTTCCACATGGATGGGCGCAGAGTCAATCACCACC 2928
Qy 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSer 60

Db 2929 AGCACCCGAACTGGGCCCTCCACCTACCAACCAACCTCTACAAACAATTTCCAGC 2988
Qy 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTyrPheGlyTyr 80
Db 2989 CAATCA--GGAGCTCGAACGACAACTACTTTGGCTACAGACCCCTTGGGGGTAT 3045
Qy 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTyrGlnArgLeuLeuAsn 100
Db 3046 TTTGACTTCAACAGATTCCACTGGCACTTTTCCACCGTGACTGGCAAGACTCATCAAC 3105
Qy 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 120
Db 3106 AACAACTGGGATTTCCGACCCAGAGACTCAACTTCAAGCTCTTAACTTCAAGTCAAA 3165
Qy 121 GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 140
Db 3166 GAGTCACGCGAGATGACGGTACGACGAGGATTCGCAATAACCTTACCAGCAGGTTTCAG 3225
Qy 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
Db 3226 GTGTTTACTGACTCGGAGTACAGTCCCGTACGTCCTCGGCTCGGCGCATCAAGGATGC 3285
Qy 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
Db 3286 CTCCCGCGTTTCCAGCAGACGCTTTCATGTGTCACAGTATGGATACCTCACCTCGAAC 3345
Qy 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
Db 3346 AACGGGAGTCAGGAGTAGGAGCTTTCATTTACTGCCTGGAGTAGTTCCTTCTTCAG 3405
Qy 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
Db 3406 ATGCTGCGTACCGGAAACAACCTTACCTTACGTACACTTTTGGAGCGTTCTCTTCCAC 3465
Qy 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
Db 3466 AGCAGTACGCTCACAGCAGAGTCTGACCGTCTCATGATTCCTCATCGACAGTAC 3525
Qy 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeu 260
Db 3526 CTGTATTACTTGAGCAGAACAACTCCAAGTGGACCAACCCAGCAGTCAAGGCTTCAG 3585
Qy 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTyrLeuProGlyPro 280
Db 3586 TTTTCTCAGCGCGGACGAGTGACATCTCGGACCACTCTAGCAACTGGCTTCTCTGGACCC 3645
Qy 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsnPheThr 300
Db 3646 TGTACCCCGCAGCAGGATATCAAGACATCTGCGGATACACACAGTGAATACTCG 3705
Qy 301 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr 320
Db 3706 TGGACTGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTGGTGAATCCGGGCCG 3765
Qy 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
Db 3766 GCCATGGCAAGCCACAAAGGACATGAAGAAAGTTTTCCTCAGAGCGGGTTCTCATC 3825
Qy 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
Db 3826 TTTGGGAAGCAGGCTCAGAGAAACAATGTGGACATTGAAAGGTCATGATTACAGAC 3885
Qy 361 GluGluGluIleLeuAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
Db 3886 GAAGAGGAAATCAGGACAACCAATCCCGTGGCTACGGAGCAGTAGTGTCTGTATCTACC 3945
Qy 381 AsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
Db 3946 AACCTCCAGAGGCAACAGACAAGCAGCTACCGCAGATGTCAACACACAAGGCGTTCTT 4005
Qy 401 ProGlyMetValTyrGlnAspArgAspValTyrLeuGlnGlyProIleTyrAlaLysIle 420

Db 4006 CCAGGCATGTCTGGCAGGACAGAGATGTGTACTCTCAGGGGCCCATCTCGGCAAGATT 4065
Qy 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
Db 4066 CCACACAGGACGACGACATTTTACCCTCTCCCTCATGGGTGGATTCCGACTTAAACAC 4125
Qy 441 ProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460
Db 4126 CTTCTCCACAGATTCTCATCAAGAACACCCCGGTACCTCGGAATCTTCCGACCCTTC 4185
Qy 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
Db 4186 AGTGGCGCAAGTTTGTCTTCTTCATCACAGTACTCCACGGGACAGGTTCAGGTGGAG 4245
Qy 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTyrAsnProGluValGlnTyrThr 500
Db 4246 ATCGAGTGGAGCTGCAGAGGAAAAACAGCAACCGTGGAAATCCGAAATTCAGTACACT 4305
Qy 501 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 520
Db 4306 TCCAACCTACAAACAGTCTGTTAATGTGGACTTTTACTGTGGACACTAATGGCGTGTATTC 4365
Qy 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
Db 4366 GAGCTCGCCCATTTGGCACCAGATACCTGACTCGTAATCTG 4407

RESULT 9
US-10-136-819-6
; Sequence 6, Application US/10136819
; Publication No. US20030166593A1
; GENERAL INFORMATION:
; APPLICANT: Chien, Kenneth
; APPLICANT: Hoshijima, Masahiko
; TITLE OF INVENTION: No. US20030166593A1-viral vesicle vector for cardiac specific ge
; FILE REFERENCE: 6627-PAL198
; CURRENT APPLICATION NUMBER: US/10/136,819
; CURRENT PILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: 60/287,423
; PRIOR FILING DATE: 2001-04-30
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 4679
; TYPE: DNA
; ORGANISM: adeno-associated virus 2
US-10-136-819-6

Alignment Scores:
Pred. No.: 7,92e-279 Length: 4679
Score: 2481.50 Matches: 445
Percent Similarity: 90.26% Conservative: 37
Best Local Similarity: 83.33% Mismatches: 51
Query Match: 85.39% Indels: 1
DB: 13 Gaps: 1

US-09-807-802A-17 (1-534) x US-10-136-819-6 (1-4679)

Qy 1 MetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGly 20
Db 2809 ATGGCTACAGCAGTGGCGCACCAATGCGACAAATAACGAGGGCGCGGAGTGGGT 2868
Qy 21 AsnAlaSerGlyAsnTyrHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 40
Db 2869 AATTCCTCGGAAATTTGGCATTTCCATGATGGCGGCGACAGTCAATCACCACC 2928
Qy 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSer 60
Db 2929 AGCACCCGACCTGGGGCTTCCGCCCTACACACCACTCTACAAACAATTTCCAGC 2988
Qy 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTyrPheGlyTyr 80
Db 2989 CAATCA--GGAGCTCGAACGACAACTACTTCTGGCTACAGCACCCCTTGGGGGTAT 3045

QY 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
DB 3046 TTGTGACTTCAACAGAGATTCCACTGCCACTTTTCCACAGCGTGGCAAGAGACTCATCAAC 3105
QY 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnVallys 120
DB 3106 AACAACTGGGGATTCCGACCCAAAGAGACTCAACTTCAAGCTCTTTAAACATCAAGTCAA 3165
QY 121 GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 140
DB 3166 GAGGTCAACAGATGACGGTACGACGAGATTTGCGCAATAACCTTACCAGCAGCGTTTCA 3225
QY 141 ValPheSerAspSerGlnTrpGlnLeuProTrpValLeuGlySerAlaHisGlnGlyCys 160
DB 3226 GTGTTTACTGACTCGGAGTACAGCTCCCGTAGCTCCCTCGGCTCGGCGATCAGATGC 3285
QY 161 LeuProProPheProAlaAspValPheMetIleProGlnTrpGlyTrpLeuThrLeuAsn 180
DB 3286 CTCGCGCGCTTCCAGCAGACGCTTTCATGGTGCCACAGTATGGATACCTCACCCCTGAAC 3345
QY 181 AsnGlySerGlnAlaValGlyArgSerSerPheTrpCysLeuLeuGlyTrpPheProSerGln 200
DB 3346 AACGGGAGTCAGCGAGTAGGAGCGCTTTCATTTTACTGCTCGAGTACTTTCCTTCTCAG 3405
QY 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTrpThrPheGluGluValProPheHis 220
DB 3406 ATGCTGCGTACCGGAACAACCTTACCTTCACTACACTTTTGGAGCGTCTCTTCCAC 3465
QY 221 SerSerTrpAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTrp 240
DB 3466 AGCAGCTACGCTCACAGCCAGAGTCTGGACCGCTCTCATGAATCCTCTCATCGACAGTAC 3525
QY 241 LeuTrpTrpLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
DB 3526 CTGTATTACTTGGAGAGAACAAACACTCCCAAGTGGAAACCAACCCAGCGAGTCAAGGCTTCA 3585
QY 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
DB 3586 TTTTCTCAGCGCGGAGTGACATTCGGGACCACTTAGGACTGGCTTCTTCTGGACCC 3645
QY 281 CysTrpArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsnPheThr 300
DB 3646 TGTATTACCGCAGCAGCGAGTATCAAGACATCTGCGGATAACAACAACAGTAGTAATCTCG 3705
QY 301 TrpThrGlyValSerLysTrpAsnLeuAsnGlyArgGluSerIleLeuAsnProGlyThr 320
DB 3706 TGGACTGGAGCTACCAAGTACCACCTCAATGGCAGAGATCTCTGTGTGAATCCGGGCCCG 3765
QY 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
DB 3766 GCCATGGCAAGCCACAGACGATGATGAAGAAAGTTTTTCTCAGAGCGGGTTCTCATC 3825
QY 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
DB 3826 TTTGGGAAGCAGGCTCAGAGAAACAAATGTGGACATTTGAAGAAGTCTATGATTACAGAC 3885
QY 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
DB 3886 GAAGAGGAATACGGAACAACCAATCCCGTGGCTACGAGAGATGTGTTCTGTATCTACC 3945
QY 381 AsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
DB 3946 AACCTCCAGAGAGCAACAGACAGCAGCTTACCGCAGATGTCAACACACAGCGCTTCT 4005
QY 401 ProGlyMetValTrpGlnAspArgAspValTrpLeuGlnGlyProIleTrpAlaLysIle 420
DB 4006 CCAGGATGTGCTGGCAGGACAGAGATGTGTACTTTCAGGGGCCCATCTGGGCAAGATT 4065
QY 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyPheGlyLeuLysAsn 440
DB 4066 CCACACAGGACCGACATTTTCAACCCCTCTCCCTCATATGGGTGGATTCGGACTTAACAC 4125
QY 441 ProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460

DB 4126 CCTCTCTCAAGATTCTTCATCAGAACACCCCGGTACTTGGCAATCCTTCGACCACTTC 4185
QY 461 SerAlaThrLysPheAlaSerPheIleThrGlnTrpSerThrGlyGlnValSerValGlu 480
DB 4186 AGTCCGGCAAGTTTGCTTCTTTCATCACACAGTACTCCACGGACAGGTCCAGGTTGGAG 4245
QY 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTrpThr 500
DB 4246 ATCCAGTGGGAGCTGCGAAGAGAAACAGCAACCGCTGGAAATCCGAAATTCAGTACT 4305
QY 501 SerAsnTrpAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTrpThr 520
DB 4306 TCCAACTTCAACAAGTCTGTTAATGTGACTTTTACTGTGGACACTAATGGCGTGTATTCA 4365
QY 521 GluProArgProIleGlyThrArgTrpLeuThrArgProLeu 534
DB 4366 GAGCCTCGCCCATTTGGCACCAGATACCTGACTCTGTAATCTG 4407
RESULT 10
US-10-038-972A-12
; Sequence 12, Application US/10038972A
; Publication No. US20020192823A1
; GENERAL INFORMATION:
; APPLICANT: J. Bartlett
; TITLE OF INVENTION: RAV VECTORS AND METHODS
; FILE REFERENCE: 28335/36996US
; CURRENT APPLICATION NUMBER: US/10/038,972A
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: US 60/260,124
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; TYPE: DNA
; ORGANISM: adeno-associated virus 2
US-10-038-972A-12
Alignment Scores:
Pred. No.: 7,92e-279 Length: 4679
Score: 2481.50 Matches: 445
Percent Similarity: 90.26% Conservative: 37
Best Local Similarity: 83.33% Mismatches: 51
Query Match: 85.39% Indels: 1
DB: 14 Gaps: 1
US-09-807-802A-17 (1-534) x US-10-038-972A-12 (1-4679)
QY 1 MetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGly 20
DB 2809 ATGGCTACAGGAGTGGCGCCACCAATGGCAGACAATAACGAGGGCGCGAGTGGGT 2868
QY 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 40
DB 2869 AATTCCTCGGAAATTTGGCATTTGGCATTTCCATGTATGGGCGGACAGAGTCAATCCACC 2928
QY 41 SerThrArgThrTrpAlaLeuProThrTrpAsnAsnHisLeuTrpLysGlnIleSerSer 60
DB 2929 AGCACCCGAACTGGGCCCTGCCACCTACAAACACCACTTACAAACAAATTTCCAGC 2988
QY 61 AlaSerThrGlyAlaSerAsnAsnHisTrpPheGlyTrpSerThrProTrpGlyTrp 80
DB 2989 CAATCA---GGAGCCTCGAACGACAATCACTACTTTGGCTACAGCACCCCTTGGGGGTAT 3045
QY 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
DB 3046 TTGTGACTTCAACAGATTCCTCACTGCGCACTTTTCCACCGTGTGGCAAGACTCATCAAC 3105
QY 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnVallys 120
DB 3106 AACAACTGGGGATTCCGACCCAAAGAGACTCAACTTCAAGCTCTTTAAACATCAAGTCAA 3165

QY 121 GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 140
DB 964 GAGGTCCACGAGATGATGGAACACGACCATCGCCAAATACCTTACCAGCAGCGTGCAG 1023
QY 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
DB 1024 GTCTTCACAGACTCTGAGTACAGCTGCCCTAGCTCTCGTTCGGTCCACAGGGCTGC 1083
QY 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
DB 1084 CTTCCGCCGTTCACAGCAGAGCTCTTCATGATTCCTCAGTACGGCTACTGACTCTGAAC 1143
QY 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
DB 1144 AATGGCAGCCAGCGGTAGACGCTTCTTCTTCTACTGCTAGAGTATTTTCCCTCTCAG 1203
QY 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
DB 1204 ATGCTCAGGACGGGAACAACATTCACCTTCAGCTACACATTTTGAAGACGTGCTTCCAC 1263
QY 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
DB 1264 AGCAGCTACGCGCACAGCCAGAGCTCTGGATCGGCTGATGAATCCTCTCATTCACCACTAC 1323
QY 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeu 260
DB 1324 CTGTATTACTGAGCAAACTCAGGTACAGTGGGAACAACGACGCAATCGAGACTGCAG 1383
QY 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
DB 1384 TTCAGCCAGCTGGGCTAGCTCCATGGCTCAGCAGGCCAATACTGGCTACCGGACCC 1443
QY 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
DB 1444 AGCTACCGCAGCAGCGGAATGTCTAAGACGGCTAATGACAACAACACAGTGAATTTGCT 1503
QY 301 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleLeuAsnProGlyThr 320
DB 1504 TGCACTGCAGCCACCAATATTAACCTGAATGGGAAGAAATCTCTGGTCAATCCGGGCC 1563
QY 321 AlaMetAlaSerHisLysAspAspGluAspLysPheProMetSerGlyValMetIle 340
DB 1564 CCAATGGCCAGTCACAAGCAGCATGAGGAAGTATTTCCCATGCAAGCAATCTCATC 1623
QY 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
DB 1624 TTTGGAAACAAGGCCACAGCACTACCAATGTGGACATTTGAATCACTGCTTTATCAGAC 1683
QY 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
DB 1684 GAGAGAGAAATCAGAACACTAATCTCTGGCTCAGAGACATATACGACAGGTGGCCACC 1743
QY 381 AsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
DB 1744 AACCATCAGAGTCAGAACACCAACAGCTCTCTATGGAAGTGTGGACAGCCAGGGAATCTTA 1803
QY 401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 420
DB 1804 CCTGGAAATGTGTGGCAGGACCGCATGTCTATCTCAAGGTCCCAATTTGGGCCAAACT 1863
QY 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
DB 1864 CCTCAGACGACGACACTTTTCACTCTCTCCGCTCATGGAGGCTTTGGACTGAACAC 1923
QY 441 ProProProGlnIleLeuLysAsnThrProValProAlaAsnProProAlaGluPhe 460
DB 1924 CTTCTCTCCCGAGATCTGATCAAAAACACACACTGTGCCAGCGAATCCCGCGACCACTTTC 1983
QY 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrClyGlnValSerValGlu 480
DB 1984 ACTCTGGAAAGTTTGTCTGTTTCACTTACCCAGTATTTCCACCGGACAGGTTCAGCGTGGAA 2043

QY 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrThr 500
DB 2044 ATAGAGTGGGAGCTGCAGAAAGAAACAGCAACGCTGGAACCCAGAAATTCAGTACACC 2103
QY 501 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 520
DB 2104 TCCAACTACACAACAGTCGGTGAATGTGGAGTTTACCGTGGAGCGCAACCGTGTATTCT 2163
QY 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
DB 2164 GAACCCCGCCTATTGGCACTCGTTACCTTACCGGAACCTTG 2205

RESULT 12
US-10-291-583-54
; Sequence 54, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alvira, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; FILE REFERENCE: Sequences and Isolating No. US20030138772A1
; CURRENT APPLICATION NUMBER: US/10291583
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 54
; LENGTH: 3123
; TYPE: DNA
; ORGANISM: new AAV serotype, clone A3.4
US-10-291-583-54

Alignment Scores:
Pred. No.: 1,34e-275 Length: 3123
Score: 2451.50 Matches: 441
Percent Similarity: 89.51% Conservative: 37
Best Local Similarity: 82.58% Mismatches: 55
Query Match: 84.36% Indels: 1
DB: 13 Gaps: 1

US-09-807-802A-17 (1-534) x US-10-291-583-54 (1-3123)

QY 1 MetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGlyValAspGlyValGly 20
DB 1444 ATGGCTTCAGCGGTGGGGCCCAATGACAGCAATACGAGCGCGGAGTGGGT 1503
QY 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 40
DB 1504 AATTCCTCGGAATTTGGCATTTCCATGATGGAGGGCGCAGAGTTATCACCACC 1563
QY 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSer 60
DB 1564 AGCACAGAACTCGGCCCTCCCACTTCAATAATACCTTACCAAGCAATCTCCAGC 1623
QY 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrTrpGlyTyr 80
DB 1624 GAATCG---GGAGCCCAACAGCAACCACTTCTCGGCTACAGCACCCCTGGGGGTAT 1680
QY 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
DB 1681 TTTGACTTTACAGATTCCACTGTCTCTCACCACGTGACTGGCAGGACTCATCAAC 1740
QY 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 120

Db 1741 AACAACTGGGATTTAGACCCAGAAACTCAATTTCAAGCTCTTCAACATCAAGTCAAG 1800
Qy 121 GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 140
Db 1801 GAGGTACGCGAGATGATGGNACCAACGACCATCCCAATAACCTTACACGACGCGTGCAG 1860
Qy 141 ValPheSerAspSerGluThrGlnLeuProThrValLeuGlySerAlaHisGlnGlyCys 160
Db 1861 GTCTTTCACAGACTCTGAGTACAGCTGCCCTACGCTCCTCGGCTCCGCTCCAGGGCTGC 1920
Qy 161 LeuProProPheProAlaAspValPheMetIleProGlnThrGlyThrLeuThrLeuAsn 180
Db 1921 CTTCCGGCGTTCCAGCAGAGCTCTTATGATTCCTCAGTACGGCTACTTGGACTCTGAAC 1980
Qy 181 AsnGlySerGlnAlaValGlyArgSerSerPheThrCysLeuGluThrPheProSerGln 200
Db 1981 AATGGCAGCCAAAGCGGTAGGAGCTTCTTCACTTCTAGTCTAGAGTATTTTCCCTCTCAG 2040
Qy 201 MetLeuArgThrGlyAsnAsnPheThrPheSerThrThrPheGluGluValProPheHis 220
Db 2041 ATGCTGAGGACGGGAAACAACCTTCACTTCACTTCTAGTCTAGAGTATTTTCCCTCTCAG 2100
Qy 221 SerSerThrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGluThr 240
Db 2101 AGCAGCTACGGCAGCAGCAGAGCTGATCGGCTGATGATCTCTCAITGACAGTAC 2160
Qy 241 LeuThrThrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
Db 2161 CTGTATTACCTGAGCAAAACCTCAGGTACAAAGTGGAAACAACCGCAATCGAGACTGCAG 2220
Qy 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnThrLeuProGlyPro 280
Db 2221 TTCAGCCAAAGTGGCGCTAGCTCCATGCTCAGCAGGCCAAAAATCGCTACCGGGACCC 2280
Qy 281 CysThrArgGlnArgValSerLysThrLysThrAspAsnAsnSerAsnPheThr 300
Db 2281 AGCTACCCAGCAGCAGGATGCTTAAGACGCTAATGACAACAACAACAGTGAATTTGCT 2340
Qy 301 TrpThrGlyAlaSerLysThrAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr 320
Db 2341 TGGACTGCAGCCACCAATATTACTGAATGGAAAGAAATCTCTGGTCAATCCCGGGCCC 2400
Qy 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
Db 2401 CCAATGGCGCAGTCAACAGGAGGATGAGGAAAGTATTTCCCATGTCAGCGGAAATCTCATC 2460
Qy 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
Db 2461 TTTGGAAAACAGGCACAGGAACTACCAATGTGGACATTTGAATCAGTGTCTATTACAGAC 2520
Qy 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
Db 2521 GAAGAAGAAATCAGAAACAACATACTCTGGCTACAGAAACAATAACGGACAGGTTGCCACC 2580
Qy 381 AsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
Db 2581 AACCATCAGAGTCAGACACACACAGCTTCCATGGAAGTGTGGACAGCAGGGAATCTTA 2640
Qy 401 ProGlyMetValTrpGlnAspArgAspValTrpLeuGlnGlyProIleThrAlaLysIle 420
Db 2641 CCTGGAATGGTGTGGCAGGACCGCATGTCTATCTTCAAGGTCCCATTTGGGCCCAAACT 2700
Qy 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
Db 2701 CCTCACAGGACGACATTTTCATCTCTCGCTCATGGGAGGCTTTGGACTGAAACAC 2760
Qy 441 ProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460
Db 2761 CCTCTCCCCAGATCCTGATCAAAACAACACCTGTGCCAGCGGAATCCCGGACCACTTTC 2820
Qy 461 SerAlaThrLysPheAlaSerPheIleThrGlnThrSerThrGlyGlnValSerValGlu 480
Db 2821 ACTCTGGAAAGTTTGCTTCGTTTCATTTACCCAGTATTCACCGGACAGGTCAGCGTGGAA 2880

Qy 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnThrThr 500
Db 2881 ATAGACTGGGAGCTGCAGAAAGAAACAGCAACAGCTGGAAACCCAGAAATTCAGTACACC 2940
Qy 501 SerAsnThrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuThrThr 520
Db 2941 TCCAACTACAACAAGTCCGCTGATGTGGAGTTTACCGTGGACCAACCGGTGTTATTCT 3000
Qy 521 GluProArgProIleGlyThrArgThrLeuThrArgProLeu 534
Db 3001 GAACCCCGCCCTATTGGCACTCGTTACCTTACCGCGAATTTG 3042

RESULT 13

US-10-291-583-56
; Sequence 56, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alivira, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; TITLE OF INVENTION: Sequences and Isolating No. US20030138772A1el Sequences Identifi
; FILE REFERENCE: UPN-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56
; LENGTH: 3122
; TYPE: DNA
; ORGANISM: new AAV serotype, clone A3.7
US-10-291-583-56

Alignment Scores:
Pred. No.: 2,29e-275 Length: 3122
Score: 2449.50 Matches: 441
Percent Similarity: 89.51% Conservative: 37
Best Local Similarity: 82.58% Mismatches: 55
Query Match: 84.29% Indels: 1
DB: 13 Gaps: 1

US-09-807-802A-17 (1-534) x US-10-291-583-56 (1-3122)

Qy 1 MetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGly 20
Db 1453 ATGGCTTCAGCGGTGGGCGACCCATGGCAGACATTAACGAAGGCGCGCAGGAGTGGT 1512
Qy 21 AsnAlaSerGlyAsnThrHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 40
Db 1513 AATTCCTCGGAAATTTGGCATTCGATTCACATGATGGCGCAGACAGATTATCACCACC 1572
Qy 41 SerThrArgThrTrpAlaLeuProThrThrAsnAsnHisLeuThrLysGlnIleSerSer 60
Db 1573 AGACAAGAAGACCTGGGCGCTCCCGCCACCTACAAATATCGCTCTACAAAGCAAAATCTCAGC 1632
Qy 61 AlaSerThrGlyAlaSerAsnAspAsnHisThrPheGlyThrSerThrProTrpGlyThr 80
Db 1633 GAATCG---GGAGCCCAACAGACACCACTACTTCGGCTACAGCACCCTCGGGGTAT 1689
Qy 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
Db 1690 TTTGACTTTAACAGATTTCCACTGTCTCTCACCACGTCGACGTCGACGACTCATCAAC 1749
Qy 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 120

Db 1750 AACAACTGGGGATTTAGACCCCAAGAACTCAATTTCAAGCTCTTCAACATCCAACTCAAG 1809
Qy 121 GluValThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 140
Db 1810 GAGGTCAACAGAAATGATGGAACACAGACCATCGCCAAATTAACCTTACACGACCGGTGAG 1869
Qy 141 ValPheSerAspSerGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
Db 1870 GTCTTCACAGACTCTGAGTACAGCTGCCCTAGCTCCTCGGTTCGGCTCACACGAGGTGC 1929
Qy 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
Db 1930 CTTCCGCCCTTCCACGACAGCTTCTCATGATTCCTCAGTACGGCTACTTGACTCTGAAC 1989
Qy 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
Db 1990 AATGGCAGCCAAACGGGTAGGAGCTTCTTCAATCTACTGTCTAGAGTATTTTCCCTCTCAG 2049
Qy 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
Db 2050 ATGCTGAGACGGGAAACAACTTCACTTCAGTACACTTTTGAAGACGTGCTTTCAC 2109
Qy 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
Db 2110 AGCAGCTACGCGCACAGCCAGAGTCTGGATCGGCTGATGAATCCTCTCATTCAGCAGTAC 2169
Qy 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
Db 2170 CTGTATTACGTGCAAAACTCAGGCTACAGTGGAAACACGCAATCGAGACTGCGAG 2229
Qy 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTyrLeuProGlyPro 280
Db 2230 TTCAGCAAGCTGGGCTAGCTCCATGGCTCAGCAGCCCAAACTGGTCTACGGGACCC 2289
Qy 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
Db 2290 AGCTACCGACAGCGCAATGTCTAAGACGGCTAATGACAAACCAACAGTGAATTTGCT 2349
Qy 301 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleLeuAsnProGlyThr 320
Db 2350 TGGAGCTGCAGCCCAACCAATATTACCTGAATGGGAAGAAATCTCTGTCAATCCCGGCCCC 2409
Qy 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
Db 2410 CCAATGGCCAGTCAACAGGACGATGAGGAAAGTATTTCCCAATGCACGGAATCTCATC 2469
Qy 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
Db 2470 TTTGGAAAACAGGACACAGCACTACCAATGTGGACATTTGAATCAGTGTCTATTACAGAC 2529
Qy 361 GluGluGluLeuLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
Db 2530 GAAGAAGAATACAGAACTAATCTGTGGCTACAGAAACATACCGACAGGTGGCCACC 2589
Qy 381 AsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
Db 2590 AACCATCAGAGTCAGAACACACACAGCTTCTATGGAAGTGTGGACACCGAGGAATCTTA 2649
Qy 401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 420
Db 2650 CCTGGAATGTGTGGCAGGACCGGATGTCTATCTTCAAGTCCCATTTGGGCCAAACT 2709
Qy 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
Db 2710 CCTCACAGGACGACACTTCTATCTCTCCGCTCATGGAGGCTTTGACCTGAACACT 2769
Qy 441 ProProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460
Db 2770 CTTCTCTCCAGATCCTGATCMAAAACACACACTGTGCGAGCAATCCCGGACCACTTTC 2829
Qy 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
:::

Db 2830 ACTCCTGAAAGTTTGTCTGTTTCATTACCCAGTATTTCCACCGACAGGTGAGTGGAA 2889
Qy 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrThr 500
Db 2890 ATAGAGTGGGAGCTGCGAAGAAAGAAACGAAACGCTTGGAAACCCAGAAATTCAGTACAC 2949
Qy 501 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 520
Db 2950 TCCAACATACAACAGTCCGTGAATGTGGAGTTTACCGTGGAGCGCAACCGTGTATTCT 3009
Qy 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
Db 3010 GAACCCCGCCCTATTTGGCACTCTGTACCTTACCCGGAACCTTG 3051
RESULT 14
US-10-291-583-55
; Sequence 55, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alvira, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; FILE REFERENCE: UPN-027350A
; CURRENT APPLICATION NUMBER: US/10/291.583
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 55
; LENGTH: 3113
; TYPE: DNA
; ORGANISM: new AAV serotype, clone A3.5
US-10-291-583-55
Alignment Scores:
Pred. No.: 6,7e-275 Length: 3113
Score: 2445.50 Matches: 440
Percent Similarity: 89.51% Conservative: 38
Best Local Similarity: 82.40% Mismatches: 55
Query Match: 84.15% Indels: 1
DB: 13 Gaps: 1
US-09-807-802A-17 (1-534) x US-10-291-583-55 (1-3113)
Qy 1 MetAlaSerGlyGlyGlyValAlaProMetAlaAspAsnAsnGlyValaAspGlyValGly 20
Db 1444 ATGGCTTCAGCGCGTGGGGCCCAATGCGACAAATAACGAAGGCGCGCAGGTGGGT 1503
Qy 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 40
Db 1504 AATTCTCGGNAATTTGGCATTTCCATGATGATGGCGACAGAGTTATCACCACC 1563
Qy 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrIleGlnIleSerSer 60
Db 1564 AGCACAGAAGAACCTGGGCCCTCCCACTTACAATAATACCTTACAAGCAAAATCTCCAGC 1623
Qy 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
Db 1624 GAATCG---GGAGCCCAACCAACCAACCACTACTTTCGGCTACAGCACCCCTTGGGGGTAT 1680
Qy 81 PheAppPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
Db 1681 TTTGACTTTAACAGATTCCACTGTCTCTTCTCACCACGCTGACTGGCAGGACTCATCAAT 1740


```
Qy 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 120
Db 1132 AACAACTGGGATTCGGGCCAGAAAGTGGGTTTCAAGTTGTTCAACATCCAGGTCAAG 1191
Qy 121 GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 140
Db 1192 GAGTCAACGACGACGCGGTACGACCATCGCATACCTTACCAGCACGATTCAG 1251
Qy 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
Db 1252 GTCTTCTCGGACTCGGAGTACCAACTGCGGTACGCTCGGCTCGGCACACGAGGCTGC 1311
Qy 161 LeuProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
Db 1312 CTCCCTCCGCTCCCTCGCGGACGTGTTCATGATTCCTCAGTACGGGATATCTGACTCTAACC 1371
Qy 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
Db 1372 AACGCGAGTCAGTCTGTGGAGGTTCTCTCTTACTGCTGGAGTACTTTCTCTTCAG 1431
Qy 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
Db 1432 ATGCTGAGAACCGGCAATAACTTTGAAATTCAGTACACCTTTGAGGAAGTGCCTTCCAC 1491
Qy 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
Db 1492 AGCACTATGCGCACAGCGAGGCTTGACCGGCTGTGATGAAATCCCTCATCGACCGTAC 1551
Qy 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
Db 1552 CTGTACTACCTGGCCCGGACCCAGAGCACTACGGGGTCC-----ACAAGGAGCTGCAG 1605
Qy 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
Db 1606 TTCCATCAGGCTGGGCCCAACACCACTGGCCGAGCAATCAAAAGAACTGGCTGCCCGGACCC 1665
Qy 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
Db 1666 TGTATTCGGCAGCAGAGACTGTCAAAAACATAGACAGCAACACACAGTAACTTTGCC 1725
Qy 301 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleAsnProGlyThr 320
Db 1726 TGGACCGGGGCCCACTAAATACCATCTGAAATGTAGAAATTCATTAACCAACCCGGCGGTA 1785
Qy 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
Db 1786 GCCATGGCCCAACCAAGGACGACGAGGACCAAGTTCTTTCCCATCAACGAGTGTGGTT 1845
Qy 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
Db 1846 TTTGGCAAAACGGGGCT---GCCAACAAGACACCGCTGGAAAACGTGCTAATGACCAGC 1902
Qy 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
Db 1903 GAGGAGAGATCAAAACCAACCAATCCCGTGGCTACAGAGAATACGGTGTGTCTCCAGC 1962
Qy 381 AsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
Db 1963 AACCTGCAATCGTCTACGCGCGGACCCAGACAGACTGTCTCAACGAGCGGGGCTCTG 2022
Qy 401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 420
Db 2023 CCGGGCATGTCTGGCAGAACCGGGACGTGTACTCGAGGGTCCCATCTGGGCCAAAATT 2082
Qy 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
Db 2083 CCTCACACGAGCGGCAACTTTCACCGCTCTCCCTGTATGGCGGATTTGGACTCAAAAC 2142
Qy 441 ProProGlnIleLeuLysAsnThrProValProAlaAsnProProAlaGluPhe 460
Db 2143 CCGCCTCCTCAATTCATCAAAAAACCCCGGTACTGCTTAATCCTCCAGAGGTGTTT 2202
```

Search completed: February 15, 2004, 02:07:54
Job time : 590.857 secs

```
Qy 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
Db 2203 ACTCTCGCCAAAGTTTGCTCATTTATCAGCAGTACAGCACCGCCAGGTCAAGGTGAG 2262
Qy 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrThr 500
Db 2263 ATCGAGTGGGAACCTGCAGAAAAGAAAACAGCAAAACGCTGGAAATCCAGAGATTCAGTACAC 2322
Qy 501 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 520
Db 2323 TCAATTTATCCCAAGTCTAATAATGTGGAATTTGCTGTCAACACGAGGAGGTTTATACT 2382
Qy 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
Db 2383 GAGCCTGCCCCCATTTGGCACCCCGTTACTCACCCGTAACCTG 2424
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: February 14, 2004, 15:19:20 ; Search time 7481.29 Seconds
(without alignments)
4024.639 Million cell updates/sec

Title: US-09-807-802A-13
Perfect score: 3989
Sequence: 1 MAADGYLPDWLEDNLSGIR.....NNGLYTEPRPIGTRYLTRLPL 736

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US9807802/runat_11022004_175607_15929/app.query.fasta_1.2389
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0 -1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-OUTFMT=ptc -NORM=score=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODES=LOCAL
-USER=US9807802@cgn_1_1 12216 @runat_11022004_175607_15929 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_on.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sta.*
28: em_uni.*

29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|------------------------|
| 1 | 3989 | 100.0 | 4718 | 14 | AF063497 Adeno-888 |
| 2 | 3963 | 99.3 | 4683 | 14 | AF028704 Adeno-888 |
| 3 | 3511 | 88.0 | 4722 | 14 | AF028705 Adeno-888 |
| 4 | 3494 | 87.6 | 4726 | 14 | AF028704 Adeno-888 |
| 5 | 3473 | 87.1 | 2211 | 14 | AY242997 Non-human |
| 6 | 3448 | 86.4 | 2217 | 14 | AY243015 Non-human |
| 7 | 3425 | 86.3 | 4721 | 14 | AF513851 Adeno-888 |
| 8 | 3439 | 86.2 | 8179 | 6 | AR264580 Sequence |
| 9 | 3438 | 86.2 | 2217 | 14 | AY243007 Non-human |
| 10 | 3434 | 86.1 | 2187 | 14 | AY243006 Non-human |
| 11 | 3432 | 86.0 | 2217 | 14 | AY243009 Non-human |
| 12 | 3429 | 86.0 | 2217 | 14 | AY243023 Non-human |
| 13 | 3424.5 | 85.8 | 2214 | 14 | AY243020 Non-human |
| 14 | 3423.5 | 85.8 | 2208 | 14 | AY243021 Non-human |
| 15 | 3422.5 | 85.8 | 2190 | 14 | AY242998 Non-human |
| 16 | 3422 | 85.8 | 2217 | 14 | AY243022 Non-human |
| 17 | 3420.5 | 85.7 | 2202 | 14 | AY243004 Non-human |
| 18 | 3420 | 85.7 | 2187 | 14 | AY243013 Non-human |
| 19 | 3417.5 | 85.7 | 2190 | 14 | AY242999 Non-human |
| 20 | 3416.5 | 85.6 | 2202 | 14 | AY243012 Non-human |
| 21 | 3414.5 | 85.6 | 2190 | 14 | AY243000 Non-human |
| 22 | 3414.5 | 85.6 | 2196 | 14 | AY243010 Non-human |
| 23 | 3413.5 | 85.6 | 2208 | 14 | AY243008 Non-human |
| 24 | 3408.5 | 85.4 | 2196 | 14 | AY243011 Non-human |
| 25 | 3406 | 85.4 | 2187 | 14 | AY243016 Non-human |
| 26 | 3405 | 85.4 | 4393 | 14 | AF513852 Adeno-888 |
| 27 | 3402.5 | 85.3 | 4072 | 6 | AR222046 Sequence |
| 28 | 3402.5 | 85.3 | 4072 | 6 | AX205074 Sequence |
| 29 | 3402.5 | 85.3 | 4679 | 6 | AX282480 Sequence |
| 30 | 3402.5 | 85.3 | 4679 | 14 | AF043303 Adeno-888 |
| 31 | 3402.5 | 85.3 | 7327 | 12 | AF369963 Cloning v |
| 32 | 3402.5 | 85.3 | 7557 | 6 | AR222045 Sequence |
| 33 | 3402.5 | 85.3 | 7557 | 6 | AX205073 Sequence |
| 34 | 3402.5 | 85.3 | 8698 | 6 | AR222044 Sequence |
| 35 | 3402.5 | 85.3 | 8698 | 6 | AX205072 Sequence |
| 36 | 3402 | 85.3 | 2187 | 14 | AY243017 Non-human |
| 37 | 3398.5 | 85.2 | 2202 | 14 | AY243014 Non-human |
| 38 | 3391 | 85.0 | 2187 | 14 | AY243018 Non-human |
| 39 | 3386 | 84.9 | 2187 | 14 | AY243019 Non-human |
| 40 | 3360.5 | 84.2 | 4680 | 6 | AR028767 Sequence 1 |
| 41 | 3360.5 | 84.2 | 4680 | 6 | AX135805 Sequence |
| 42 | 3322.5 | 83.3 | 4675 | 6 | AX135805 Sequence |
| 43 | 3322.5 | 83.3 | 4675 | 6 | AX286292 Sequence |
| 44 | 3322.5 | 83.3 | 4675 | 6 | BD094552 Method of |
| 45 | 3322.5 | 83.3 | 4675 | 14 | AA2CG J01901 Adeno-888 |

ALIGNMENTS

AF063497 4718 bp DNA linear VZL 27-APR-1999
LOCUS Adeno-associated virus 1, complete genome.
DEFINITION AF063497
ACCESSION AF063497
VERSION AF063497.1 GI:4689096
KEYWORDS
SOURCE
ORGANISM
Adeno-associated virus 1
Adeno-associated virus 1
Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
REFERENCE 1 (bases 1 to 4718)
Xiao, W., Chirmule, N., Berta, S.C., McCullough, B., Gao, G. and
Wilson, J.M.
Gene therapy vectors based on adeno-associated virus type 1
1 (bases 1 to 4718)
J. Virol. 73 (5), 3994-4003 (1999)
MEDLINE 99214338
PUBMED 10196295
REFERENCE 2 (bases 1 to 4718)
Xiao, W. and Wilson, J.M.
Direct Submission
TITLES Submitted (05-MAY-1998) IHGT, University of Pennsylvania, 3601
Spruce Street, Philadelphia, PA 19104, USA
JOURNAL Location/Qualifiers
FEATURES
source
1..4718
/organism="Adeno-associated virus 1"
/mol_type="genomic DNA"
/db_xref="taxon:85106"
/note="AAV1"
335..2206
/codon_start=1
/product="nonstructural protein"
/protein_id="AAD27758.1"
/db_xref="GI:4689098"
translations="MPSFEIVIKVPSDLDELHLPGLSDSFVSWAEKEWELPPDSMD
LNLLEQAPITVAEKLDLDELHLPGLSDSFVSWAEKEWELPPDSMD
MVLRLSQRDLKLVITIRGIEPLTPNPAVTKTRNGAGGKGVVDECIENILLPK
TOPELQAWNTMEETYSACINLAERKELVAHLTHVSTOQONKENLNPNSDAPVRS
KTSARYMELVGLVDRTITSEKQWIOEDQASYISFNASNSRSQIKALDNDAGKIMAL
TKSAPYLVPAPADIKTNRIYRIELNGEYPAVAGSVFLGWAOKRFGKRTNITWLF
PATTGNTIAEAHAAPVFCVNTNENPPENDCVDMVWEEGKMTAKVVESSAKA
ILGSKRVDDKCKSAQIDPTPIVITSNTNCAVIDGNSITFEHQOPLQDRMKEPEL
TRLEHDFGKTKQEVKPRFQADHVEAFYRKGANKRPAPDADKSEPKRA
CPVADPSTSDAEGAVDFADRYONKCSRHAGLQMLFPCKTCERNWQNFICFTHGT
RDCSECFPGVSGSPVVRKTRTKYLCAIHLLGRAPEIACSDLVNVDLDCVSEQ"
2223..4433
/codon_start=1
/product="capsid protein"
/protein_id="AAD27757.1"
/db_xref="GI:4689097"
translations="MAADGYLPDLEDNLSEGIREWDLKPGAPKPAKQKQDDGRG
LVLPGYKLPFGNLDKGEVNAADAALAEHDKAYDQQLKAGDNPYLRYNHADAEPQ
RLQEDTSFGNIGRAVFOAKRVLEPLGLVEEAKTAPGKRPEVQSPDSSSGIG
KTQQPAKELNFGQDSDSPDPOPLGEPATPAAGVPTTWSAGGAPMADNNEGA
DVGASGNHCDSTWLDGRVITSTRTHALPTYNHLYKQISSASTGASNDNHFGY
STPWGYDFNRHFCHSPRDWQRLNNNNRFRKLNFLNFQVKEVTNDGVTITIA
NNLTSTVDFSDSEYQLPYLGAHQCLPPFPADPFMTQYGLYLTNNQSQVGRSS
FYCLEYPPSQMLATGNFTSYTFEEVPHSSYASQSLDRLANPLIDQLYLYLNRTQ
NQSAGNKLDSRSPGAGSVQPKWLPFCYRQORVSKTKTDNNNSNFTWTGASK
YNLNGESIINCPGTMAHSHKDDKFPNMGVMI FGESAGASNTALDNMTIDEBEI
KATNPATERFGTVANFQSSDTPATGDVHANGALPGWMDRDVYLOGPIWAKIPH
TDGHFPHSPGFLGNPPQLIKNTIPVPPAFEFSAFKFASFIITQSTQGVSE
IEWELQKENSKRNVPEQVTSYAKSANVDFTVDNNGLYTEPRPIGRYLRPL"

BASE COUNT 1121 a 1393 c 1273 g 931 t
ORIGIN
Alignment Scores:
Pred. No.: 2,14e-240 Length: 4718
Score: 3989.00 Matches: 736
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0
US-09-807-802A-13 (1-736) x AF063497 (1-4718)

1 MetAlaAlaAspGlyTyrLeuProAspTrpLeuGluAsnLeuSerGluGlyLeuArg 20
2223 ATGGCTGCCGATGGTATCTTCAGATTGGCTCAGGACCAACCTCTCTGAGGGCATTCGC 2282
21 GluTrpTrpAspLeuLysProGlyValaProLysProLysAlaAsnGlnGlnLysGlnAsp 40
2283 GAGTGGTGGGACTTGAACCTGGAGCCCGAAGCCCAAGCCCAACAGCAAAAGCAGGAC 2342
41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
2343 GACGCCCGGGGTCTGGTCTTCTCGGTACAAAGTACCTCGGACCCCTTCAACGGACTCGAC 2402
61 LysGlyGluProValAsnAlaAlaAspAlaAlaAlaLeuGluHisAspLysAlaTyrAsp 80
2403 AAGGGGAGCCGCTCAACGGCGGACGCGGGGCCCTCGAGCAGCAAGGSCCTACGAC 2462
81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
2463 CAGCAGCTCAAAAGCGGGTGACAATCCGTACCTCGGTATTAACCAACGCGGACGCGGATTT 2522
101 GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAsnLeuGluGlyValaValPheGln 120
2523 CAGGAGCGTCTGCAAGAAAGATACGCTTTTGGGGGCAACCTCGGGCGGACGACTCTTCAG 2582
121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGluGlyValaValPheGln 140
2583 GCCAAGAGCGGGTCTCGAACCTCTCGGTCTGGTTAGGAAGGCGCTAAGACGGCTCCT 2642
141 GlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSerGlyLleGly 160
2643 GGAAGAAACCGTCCGCTAGAGCAGTCCGACCAAGAGCCAGACTCTCTCTCGGGCATTCGC 2702
161 LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
2703 AAGACAGCGCAGACCGCCGCTAAAGAGACTCAATTTTGGTCACTGGCGACTTCAGAG 2762
181 SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro 200
2763 TCAGTCCCGCATCCACACCTCTCGGAGAACCTCCAGCAACCCCGCTGCTGTGGACCT 2822
201 ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGly 220
2823 ACTACATGGCTTCAGCGGTGGCGCCCAATGGCAGACAATAACGAAGGCCCGCAGCGA 2882
221 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIle 240
2883 GTGGGTAAATGCTTCAAGAAATTTGGCATTTCCGATTTCCACATGCTGGCGACAGACTC 2942
241 ThrThrSerThrArgThrTrpAlaLeuProThrThrThrThrThrThrThrThrThr 260
2943 ACCACCGACCGCCGACCTGGGCTTGGCCACCTTACCAATTAACCACTCTTACAGCAAAATC 3002
261 SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrp 280
3003 TCCAGTGTCTCAACGGGGGCCAGCAACCAACCACTACTCTCGGCTACAGACCCCTCG 3062
281 GlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu 300
3063 GGGTATTTGATTTCAACAGATTTCCACTGCCACTTTTACCACGCTGACTGGCAGCGACTC 3122
301 IleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGln 320
3123 ATCAACCAAAATTTGGGGATTCCGGCCCAAGAGACTCAACTTCAAACTCTTCAACATCCAA 3182
321 ValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThr 340
3183 GTCAGGAGGTCAACGACGAATGATGGCGTCACAAACCATTCGTATTAACCTTACCAGCAG 3242
341 ValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGln 360
3243 GTTCAAGTCTTCTCGGACTCGGAGTACCAAGCTTCCGTACGTCTCTCGGCTCTGCGCACCAG 3302

361 GlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr 380
3303 GGCTGCTCCCTCGTCCGGCGACGTGTTATGATTCGCAATACGGCTACCTGACG 3362
381 LeuAsnGlnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 400
3363 CTCACCAATGGCAGCCGAGCGGTGGGACGTTCATCTTTTACTGCTCGGAAATATTCCT 3422
401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 420
3423 TCTCAGATGCTGAGAACGGGCAACACTTTACCTTCAGCTACACCTTTGAGGAGTGCCT 3482
421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
3483 TTCCACAGCAGCTACGGCGCACAGCGAGCGCTGACCGGTGATGAATCTCTCATCGAC 3542
441 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAsp 460
3543 CAATACCTGTATTACCTGAACAGAACTCAAAATCAGTCCGGAAGTGCCTCAAAACAAGGAC 3602
461 LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTyrLeuPro 480
3603 TTGCTGTTAGCTGGGTCTCCAGCTGGCATGCTCTTTCAGCCCAAACTGGCTACCT 3662
481 GlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsn 500
3663 GGACCTGTTATCGCAGCAGCGGCTTCTAAAACAAAACAGACAAACAACAGCAAT 3722
501 PheThrTyrThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnPro 520
3723 TTTACCTGGAGCTGGTGTCTCAAAATATAACCTCAATGGGCGTGAATCCATCATCAACCT 3782
521 GlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyVal 540
3783 GGCACTGCTATGGCTCACACAAAGACGACGACGACGATCTCTTCCATGAGCGGTGTC 3842
541 MetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIle 560
3843 ATGATTTTGGAAAGAGAGCGCGGAGCTTCAACACATCGCATTTGGACAAATGTCATGAT 3902
561 ThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrVal 580
3903 ACAGCAGAGAGGAAATTAAGCCCACTAACCTGTGGCCACCGAAGATTTGGAGCCGTG 3962
581 AlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGly 600
3963 GCAGTCAATTTCCAGAGCAGCAGCAGCAGCCTCGCAGCGAGATGTGCATGCTATGGGA 4022
601 AlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAla 620
4023 GCATTACCTGGCATGGTGTGGCAAGATAGACGCTGTACCTGCAGCGGTCCCATTTGGGCC 4082
621 LysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeu 640
4083 AAAATTCCTCACAGATGGACATTTACCCCGCTCTCTCTATGGCGCGCTTTGGACTC 4142
641 LysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAla 660
4143 AAGAACCCGCTCTCTCAGATCTCTCATCAAAACAGCGCTGTCTCTCGCAATCTCCCGCG 4202
661 GluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSer 680
4203 GAGTTTTTCAGCTACAAAGTTTGTCTTATCATCCCAATCACTCCACAGGACCAAGTGA 4262
681 ValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTyrAsnProGluValGln 700
4263 GTGGAAATTTGAATGGGAGCTCGAAGAAACAGCAAGCGCTCGAATCCGGAAGTGCAG 4322
701 TyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeu 720
4323 TACACATCCATTTATGCAAAATCTGCCACGTTGATTTTACTGTGGACCAACATGAGCTT 4382
721 TyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736

||||| 4383 TATAGTACGCTCGCCCATTCGACCCCTTACCTTACCGCTCCCTG 4430
RESULT 2
AF028704
LOCUS Adeno-associated virus 6, complete genome. VRL 12-JAN-1998
DEFINITION AF028704
ACCESSION AF028704
VERSION AF028704.1 GI:2766605
KEYWORDS
SOURCE Adeno-associated virus 6
ORGANISM Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
REFERENCE 1 (bases 1 to 4683)
AUTHORS Rutledge, E.A., Halbert, C.L. and Russell, D.W.
TITLE Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2
J. Virol. 72 (1), 309-319 (1998)
JOURNAL 98080418
MEDLINE 9420229
PUBMED
REFERENCE 2 (bases 1 to 4683)
AUTHORS Rutledge, E.A. and Russell, D.W.
TITLE Direct Submission
JOURNAL Submitted (24-SEP-1997) Hematology, Univ. of Washington, Box 357720, Seattle, WA 98195, USA
FEATURES
Location/Qualifiers
1. .4683
/organism="Adeno-associated virus 6"
/mol_type="genomic DNA"
/db_xref="taxon:68558"
/note="AAV6"
320. .2191
/codon_start=1
/evidence=not experimental
/product="nonstructural protein Rep78"
/protein_id="AAB95449.1"
/db_xref="GI:2766606"
2208. .4418
/codon_start=1
/evidence=not experimental
/product="capsid protein VP1"
/protein_id="AAB95450.1"
/db_xref="GI:2766607"
/translation="MAADQGLPDMLEDNISEGIREWMDLKGAPKPKANQQKQDDGRG
LVLPGYKYLGFNGLDKGPNVNAADAAALHDKAYDQQLKAGDNPLRYNHADAEPQ
RLEQDTSFGNGLGRAVFAKRVLPGLVBEGAKTAPGKTPVEGSPQEPDSSSISG
KTQOPAKRLNFGTGDSEVPDPOPLGEPATPAAGPTTMAAGCGAPMADNNEGA
DVGASGNWNCDSWLGDRVATTSRTWALPTYNHLYKOISSASTGSNDNHYFGY
STPWGTFDRFHCHFSRDMQRLNNNNGFRPKRLNFKLNIQVKEVTNDQVTTIA
NNLTSTVQFSSEYQLPYVLGSAHQGLPPFPADVFMIPQYGLTLNNGSQVGRSS
FYCLEYFPQMLRTGNNFTSYTFEDVPFHSYASHSQSLDRMLNPLIDQLYLYLNRQ
NCSGQNKKDLFSRSGPAGMSVQPKWLPFGCYQRQVRVSKTKTNNNSNFTWTGAK
YNLNGRESINFTAMASHKDDKFFPMGSGVMI FGESAGASNTALDNVMI TDEEII
KATNPVATPERFCTVAVNLQSSDTPATGTVHVMGALPGVMQDRDVLVLOGPIWAKLPH
TQGHFHPSLMGFGGLKHPPOQLIKNTVPANPPAFSAFSPASITQYSTQGVSE
IWEVQKENSKNWNPVQTSNYAKSANDVFTVDNNGLYTEPRPIGTRYLTRPL"

BASE COUNT 1114 a 1363 c 1277 g 929 t
ORIGIN
Alignment Scores: 9.01e-239 Length: 4683
Pred. No.: 3963.00 Matches: 730
Score:

| | | | |
|---|--|--|------|
| Percent Similarity: | 99.59% | Conservative: | 3 |
| Best Local Similarity: | 99.18% | Mismatches: | 3 |
| Query Match: | 99.35% | Indels: | 0 |
| DB: | 14 | Gaps: | 0 |
| US-09-807-802A-13 (1-736) x AF028704 (1-4683) | | | |
| Qy | 1 | MetaAlaAspGlyTyrLeuProAspTrpLeuGluAspAenLeuSerGluGlyIleArg | 20 |
| Db | 2208 | ATGGCTCCGATGGTTATCTTCAGATTGGCTCGAGGACAACTCTCTGAGGGCAATTGCG | 2267 |
| Qy | 21 | GluTrpTrpAspLeuLysProGlyAlaProLysProLysAlaAenGlnGlnLysGlnAsp | 40 |
| Db | 2268 | GAGTGGTGGGACTTGAACCTGGAGCCCGAAACCCAAAGCAACACCAAGCAAGAGCAGAC | 2327 |
| Qy | 41 | AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAenGlyLeuAsp | 60 |
| Db | 2328 | GACGGCTGGGTCTGGTCTCTCGCTACAGTACTCTCGGACCTTCAACGGACTCGAC | 2387 |
| Qy | 61 | LysGlyGluProValAenAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp | 80 |
| Db | 2388 | AAGGGGAGCCGCTCAACGGCGGTGTCAGCGCCCTCGAGCACGACAAAGGCTACGAC | 2447 |
| Qy | 81 | GlnGlnLeuLysAlaGlyAspAenProTyrLeuArgTyrAenHisAlaAspAlaGluPhe | 100 |
| Db | 2448 | CAGCAGCTCAAGCGGGTGCAATCCGTACTCGGTATTAACACCGCCGACGCGAGTTT | 2507 |
| Qy | 101 | GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAenLeuGlyArgAlaValPheGln | 120 |
| Db | 2508 | CAGAGCGTCTGCAAGAGATACGTCTTTTGGGGCAACCTCGGGCGGANGTCTTCAG | 2567 |
| Qy | 121 | AlaLysIysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaPro | 140 |
| Db | 2568 | GCCAAGAGGGTCTCGAACCTTTTGGTCTGGTTGAGGAAGTGTCTAAGACGGCTCT | 2627 |
| Qy | 141 | GlyLysIysArgProValGluGlnSerProGlnGluProAspSerSerSerGlyIleGly | 160 |
| Db | 2628 | GGNAAGAAACCTCCGGTACAGCAGTCGCCCAAGAGCCAGACTCTCTCTCGGGCATTTGGC | 2687 |
| Qy | 161 | LysThrGlyGlnGlnProAlaLysLysArgLeuAenPheGlyGlnThrGlyAspSerGlu | 180 |
| Db | 2698 | ANGACGCCAGCAGCCCGCTTAAAGAGACTCAATTTGTCTAGACTGGCGACTCAGAG | 2747 |
| Qy | 181 | SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro | 200 |
| Db | 2748 | TCAGTCCCGACCCACAACTCTCGAGAACTTCAGCAACCTCCAGCAACCCCGCTGCTGTGGACCT | 2807 |
| Qy | 201 | ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAenAenGluGlyAlaAspGly | 220 |
| Db | 2808 | ACTACAATGGCTTCAGCGGGTGGCGCACCAATGGCAGACAATAACGAAGGCCCGCAGCGGA | 2867 |
| Qy | 221 | ValGlyAenAlaSerGlyAenTrpHisCysAspSerThrTrpLeuGlyAspArgValile | 240 |
| Db | 2868 | GTGGGTATGGCTCAGGAATTTGGATTCGATTCACATGCTGGCGCAGAGTCAATC | 2927 |
| Qy | 241 | ThrThrSerThrArgThrTrpAlaLeuProThrTyrAenAenHisLeuTyrLysGlnIle | 260 |
| Db | 2928 | ACCACGACCCGCAACATGGGCTTGGCCCTTACCACTATTAACAACCACTCTACAAGCAATC | 2987 |
| Qy | 261 | SerSerAlaSerThrGlyAlaSerAenAenHisTyrPheGlyTyrSerThrProTrp | 280 |
| Db | 2988 | TCCAGTGTCTCAACGGGGCCGACCAACCACTACTCTCGGCTACAGCACCCCTGG | 3047 |
| Qy | 281 | GlyTyrPheAspPheAenArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu | 300 |
| Db | 3048 | GGGTATTTGATTTCAACAGATTCCACTGCCATTTCTCACCAGCTGACTGGCAGGACTC | 3107 |
| Qy | 301 | IleAenAenAsnTrpGlyPheArgProLysArgLeuAenPheLysLeuPheAenIleGln | 320 |
| Db | 3108 | ATCAACAACAAATTTGGGATTCGGGCCCAAGAGACTCAACTTCAAGCTCTTCAACATCCA | 3167 |
| Qy | 321 | ValLysGluValThrThrAenAspGlyValThrThrIleAlaAenAenLeuThrSerThr | 340 |
| Db | | | |
| 3168 | GTCAAGAGGGTCAACGAGAAATGATGGCTGTCAGCACCATCGCTAATAACCTTACCAGCAGC | 3227 | Db |
| 341 | ValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGln | 360 | Qy |
| 3228 | GTTCAAGTCTTCTCGGACTCGGAGTACCAGTTGCCGTAGCTCTCGGCTCTGGGCAACCAG | 3287 | Db |
| 361 | GlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr | 380 | Qy |
| 3288 | GGCTGCTCCCTCCGTCCCGCGGACGTGTTTCATGATTCGCGAGTACGGCTACCTAAACG | 3347 | Db |
| 381 | LeuAenAenGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro | 400 | Qy |
| 3348 | CTCAACAATGGCAGCCAGGAGTGGGACGGTCATCTTTTACTGCTCGGAATATTTCCCA | 3407 | Db |
| 401 | SerGlnMetLeuArgThrGlyAenAenPheThrPheSerTyrThrPheGluGluValPro | 420 | Qy |
| 3408 | TCGCAGATGCTGAGAACGGGCAATAACTTTACCTTCAGCTACACCTTCGAGGACGTGCCT | 3467 | Db |
| 421 | PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAenProLeuIleAsp | 440 | Qy |
| 3468 | TTCCACAGCAGCTACCGCACAGCCAGAGCTGGACCGGCTGATGAATCTCTCTCATCGAC | 3527 | Db |
| 441 | GlnTyrLeuTyrTyrLeuAenArgThrGlnAenGlnSerGlySerAlaGlnAenLysAsp | 460 | Qy |
| 3528 | CAGTACTGTATTTACCTGAACAGACTCAGAACTCAGTCCGGAAGTCCCAAAAACAGGAC | 3587 | Db |
| 461 | LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAenTrpLeuPro | 480 | Qy |
| 3588 | TTGCTGTTTAGCCGGGGTCTCCAGCTGGCATGTCTGTTTCAGCCCAAAAACCTGGCTACCT | 3647 | Db |
| 481 | GlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAenAenSerAen | 500 | Qy |
| 3648 | GGACCTGTTTACCGGAGCAGCGCGTTTCTTAAACAACAAAACAGCAACAAACAGCAAC | 3707 | Db |
| 501 | PheThrTrpThrGlyAlaSerLysTyrAenLeuAenGlyArgGluSerIleIleAenPro | 520 | Qy |
| 3708 | TTTACCTGGACTGGTGGCTTCAAAATATAACTTTAATGGGGTGGATCTATATCAACCTT | 3767 | Db |
| 521 | GlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyVal | 540 | Qy |
| 3768 | GGCACTGTATGGCTTACACAAAGACGACAAAGACAAAGTTCTTCCCATGAGCGGTGTC | 3827 | Db |
| 541 | MetIlePheGlyLysGlnSerAlaGlyAlaSerAenThrAlaLeuAspAenValMetIle | 560 | Qy |
| 3828 | ATGATTTTGGAAAGAGAGCGCGGAGCTTCAACACTGCATTTGGACAAATGTCATGATC | 3887 | Db |
| 561 | ThrAspGluGluGluIleLysAlaThrAenProValAlaThrGluArgPheGlyThrVal | 580 | Qy |
| 3888 | ACAGACGAGAGGAAATCAAGCACTTAACCCCTGGCCACCCGAAAGATTTGGGACTGTG | 3947 | Db |
| 581 | AlaValAenPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGly | 600 | Qy |
| 3948 | GCAGTCAATCTCCAGAGCAGCAGACCTTCGACCGAGATGTGCATGTTATGGGA | 4007 | Db |
| 601 | AlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAla | 620 | Qy |
| 4008 | GCCTTACCTCGAATGGTGGCAAGACAGAGACTATACCTGCGGGTCTCTATTTGGGCC | 4067 | Db |
| 621 | LysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeu | 640 | Qy |
| 4068 | AAATTTCTCACCGGATGGACACTTTCACCCCTCTCTCTCATGGCGGCTTTGGACTT | 4127 | Db |
| 641 | LysAenProProGlnIleLeuIleLysAenThrProValProAlaAenProProAla | 660 | Qy |
| 4128 | AAGCACCGGCTCTCAGATCTCTCATCAAAAACACGCTGTTCTCTGCAATCTCTCCGCA | 4187 | Db |
| 661 | GluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSer | 680 | Qy |
| 4188 | GAGTTTTTGGGCTCAAAAGTTTGTCTTATTCATCCCGAGTATTTCCACAGGACAGTGGC | 4247 | Db |
| 681 | ValGluIleGluTrpGluLeuGlnLysGluAenSerLysArgTrpAenProGluValGln | 700 | Qy |
| 4248 | GTGGAGATTGATGGGAGCTGCGAAGAAAGAAAACAGCAACGCTGGATCTCCGAGTGCAG | 4307 | Db |

```

Qy 701 TyrThrSerAenTyAlaLysSerAlaAenValaAspPheThrValaAspAenAenGlyLeu 720
Db 4308 TATACATCTAATCTGCAAAATCTGCCAACTGTTGATTCTGCTGAGGACCAATAGGACTT 4367

Qy 721 TyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 4368 TATACTAGCCTGCCCATTTGGCACCGGTACCTCACCCGTCCTCCCTG 4415

RESULT 3
AF028705 4722 bp DNA linear VRL 12-JAN-1998
LOCUS
DEFINITION Adeno-associated virus 3B, complete genome.
ACCESSION AF028705
VERSION AF028705.1 GI:2766608
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 4722)
Rutledge,S.A., Halbert,C.L. and Russell,D.W.
Infectious clones and vectors derived from adeno-associated virus
(AAV) serotypes other than AAV type 2
J. Virol. 72 (1), 309-319 (1998)
98080418
9420229
2 (bases 1 to 4722)
Rutledge,S.A. and Russell,D.W.
Direct Submission
Submitted (25-SEP-1997) Hematology, Univ of Washington, Box 357720,
Seattle, WA 98195, USA
LOCATION/Qualifiers
1. 4722
/organism="Adeno-associated virus 3B"
/mol_type="genomic DNA"
/db_xref="taxon:68742"
/ncfe="AAV3B"
317..2191
/codon_start=1
/product="nonstructural protein Rep78"
/protein_id="AAB95451.1"
/db_xref="GI:2766609"
/translation="MPGFYEIVLVKPSDLDLHPLGISFNVAEKEWELPPDSMD
PNLEQAPLTVAEKQLREFVWRVRSKAPLEAFFVQFEKGETYFHLHLVETIGVKS
MVGVYQIKELKLVTRIYRGVEQLPNFVAVTKTRGAGGKGVDDCYIPNLYLPK
TQPELQAWTMDQYLSACLAEKRLVAQLHTVSGTQSEKQENQPNSDAPVIRS
KTSARYMELVGLVDRIETSEKOWIQEDQASYISFNAASNSRSQIKALDNASKIMSL
KTADPYLVGNSNPEDITKNRIYOILELNGVDPOYAAASVFLGWAKKFGKRNITLPG
PATGKTNIAEIAHAFVYGCUNWTHNEFPNDVDMVWEEGKMTAKVYESAKA
ILGSKVRVDCKKSSAQIEFPVIVTSNTNMCVADIGNSTTFEHOQLQDRMPKFL
TRRLDHPFGKVTQGVKDFRWASDHTVDVAHEFYVRGGAKRKPSANDADVSEPRQ
CTSLAQPTSDAEPADYADRYQNKSRHVGMLMLFPCKTCERMNOISNVCFTHGOR
DCGCECFPMSESQPVSVVKKTYQKLCPIHILGRAPEIACACDLANVLDLDCVSBQ
"
2208..4418
/codon_start=1
/product="capsid protein vp1"
/protein_id="AAB95452.1"
/db_xref="GI:2766610"
/translation="MAADGVLPLDLEDNLSIGREWWALPGVPOPKANQOHDNRGG
LVLPQYKVLGPNGLDGEPEVNEADAALHDKAYDQOLKAGDNPYLKYNDAEFOE
RLQDTSFGLNLRGAVQAKRKLLEPLGLVEEAATKAPGKRFPVDSQPQSDSSGVG
KSGQAPARKLNFQGTGSESVDPDPLGPEFAAPTSLGNTMVSQSGAPMADNNEGA
DVGNSGNSNHCDSQWLGDRIYTTSTRWALPTYNHLYKQISSQSGASNDNHFYGS
TPWGYFDFNRFCHFSPRDMQRLNNNNWFRPKLSFKLNIQVKEVTQNDGTTIAN
NLTVQVDFTSYQVLPVLSAQHQCGLPPPPVFWPOYVQVTLNNGSQAQVRSFP
YCLFVFPQMLRTGNPFQSVTFPDPVPHSSVAHSQSLDLRLNPLIDLYLYLNRTOG
TTSSTNQSRLLFFQAGQPQSMQLQARNWLPFCYRQRLSTANDNNNSFPWTAASK
YHLNRLSLNPGPMASHKODESEKFPFMHGNLIFGKEGTTASNAELDNVMTDEEII
RTTNPAVTEQYGTWANNLQSSNTAPTRTNVDQGLFCMVWQDRDVLQGPFWAKIPH
TDGHFHSPLMGFLGHPPIQIMIKNTFPVNPPTTFSAPKASFITQYSTGVSVSE
IEWELOKNSKRWNPICQYTSNKNYSVNVDFTVDTNGVYSEPRPIGTRTNL"
1204 a 1297 c 1243 g 978 t
BASE COUNT

```

ORIGIN

```

Alignment Scores:
Pred. No.: 1.88e-210 Length: 4722
Score: 3511.00 Matches: 640
Percent Similarity: 91.99% Conservative: 38
Best Local Similarity: 86.84% Mismatches: 57
Query Match: 88.02% Indels: 2
DB: 14 Gaps: 2

US-09-807-802A-13 (1-736) x AF028705 (1-4722)

Qy 1 MetAlaAlaAspGlyTyrLeuProAspTyrLeuGluAspAenLeuSerGluGlyValArg 20
Db 2208 ATGGCTGCTGACGGTTATCTTCAGATTGGCTCGAGGACAACCTTTCTGAAGGCATTCGT 2267

Qy 21 GluTyrTyrAspLeuLysProGlyAlaProLysProLysAlaAenGlnGlnLysGlnAsp 40
Db 2268 GAGTGGTGGCTCTGAACCTGGAGTCCCTCAACCCCAAGCAACCAACCAACGACGAC 2327

Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAenGlyLeuAsp 60
Db 2328 AACCGTCGGGGTCTTGTGCTTCGGGTTACAATACCTCGGACCCCGGTAAACGACTCGAC 2387

Qy 61 LysGlyGluProValAenAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
Db 2388 AAAGGAGAGCCGGTCAACGAGGCGGACCGCGGACCCCTCGAACAACGAAAGCTTACGAC 2447

Qy 81 GlnGlnLeuLysAlaGlyAspAenProTyrLeuArgTyrAenHisAlaAspAlaGluPhe 100
Db 2448 CAGAGCTCAAGCCGGTGACCAACCGTACTCAAGTACCAACCAACCGCGGACCGGAGTTT 2507

Qy 101 GlnGluArgLeuGlnGluAspThrSerPheGlyLysLeuGlyArgAlaValPheGln 120
Db 2508 CAGAGCGCTCTCAAGAAGATACGCTCTTTGGGGCAACCTTGGCAGAGCAGTCTTCAG 2567

Qy 121 AlaIysLysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaPro 140
Db 2568 GCCAAAAGAGAGTCTTGGCTCTGGTGTAGGAAGACGAGCTAAACCGGCTCT 2627

Qy 141 GlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSerGlyValGly 160
Db 2628 GGAAGAGAGAGGCTGTAGATCAGTCTCTCAGGAACCGGACTCATCATCTGTTGTTGGC 2687

Qy 161 LysThrGlnGlnProAlaLysLysArgLeuAenPheGlyGlnThrGlyAspSerGlu 180
Db 2688 AAATCGGGCAACACGCTGCGAGAAAAGACTAAATTTCCGTTCAGACTGCGCAGCTCAGAG 2747

Qy 181 SerValProAspProGlnProLeuGlyGluProAlaThrProAlaAlaValGlyPro 200
Db 2748 TCAGTCCCAGACCTCAACCTCTCGGAGAACCCAGCAGCCGCCCAAGTTTGGGATCT 2807

Qy 201 ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAenAenGluGlyAlaAspGly 220
Db 2808 AATACAATGGCTTCAGGCGTGGCGCACCAATGGCAGACCAATACGAGGGTGGCGATGGA 2867

Qy 221 ValGlyAenAlaSerGlyAenThrHisCysAspSerThrTrpLeuGlyAspArgValIle 240
Db 2868 GTGGTAATTCCTCAGGAATTTGGCATTTCCCAATGGGTGGCGGACGACAGTCAATC 2927

Qy 241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAenAenHisLeuTyrLysGlnIle 260
Db 2928 ACACACGACGACGACCACTGGGCGCTGCCCCACCTTACCAACCAACCAATCTCTACACCAATC 2987

Qy 261 SerSerAlaSerThrGlyAlaSerAsnAspAenHisTyrPheGlyTyrSerThrProTrp 280
Db 2988 TCCAGCCCAATCA--GGAGCTTCAACAGCAACCACTACTTTGGCTACAGCACCCCTTGG 3044

Qy 281 GlyTyrPheAspPheAenArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu 300
Db 3045 GGGTATTTGACTTTAAGATTCACATCCACTGCGCACTTCTCACCACGTCGACTGCGCAGCTC 3104

Qy 301 IleAsnAenAenTrpGlyPheArgProLysArgLeuAenPheLysLeuPheAenIleGln 320

```

```

|||||
3105 ATTAACAACAACCTGGGGATTCGGGCCCAAGAACTCAGCTTCAAGCTCTTCAACATCCAA 3164
Qy |||||||
3321 ValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThr 340
Db |||||||
3165 GTTAAAGAGGTCAACGACAGATGGCAGCAGCTATTGGCCTAATAACCTTACCAGCAGC 3224
Qy |||||||
3341 ValGlnValPheSerAspSerGluThrGlnLeuProTyrValLeuGlySerAlaHisGln 360
Db |||||||
3225 GTTCAAGTGTATACGACTCGGAGTATCAGCTCCCGTACGTGCTCGGCTCGGCGCACCA 3284
Qy |||||||
3361 GlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr 380
Db |||||||
3285 GGCCTGCTCCCGCGCTTCCAGCGAGCGTCTTCATGGTCCCTCAGTATGGATACCTCACC 3344
Qy |||||||
3381 LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 400
Db |||||||
3345 CTGAACAACGGAAGTCAACGCGTGGGACGCTCATCTTTTACTGCTCGAGTACTTCCT 3404
Qy |||||||
3401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 420
Db |||||||
3405 TCGCAGATGCTAAGACGCTGGAATTAACCTTCCAATTCAGCTATACCTTCGAGATGACT 3464
Qy |||||||
3421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
Db |||||||
3465 TTTACAGCAGCTACGCTCACAGCCAGAGTTGGATCGCTTGATGAATCCTCTTATTGAT 3524
Qy |||||||
3441 GlnTyrLeuTyrTyrLeuAsnArgThrGln---AsnGlnSerGlySerAlaGlnAsnLys 459
Db |||||||
3525 CAGTATCTGTACTCTGAACAGACGCAAGCAACCTCTCGGAACCAACCAACCAATCA 3584
Qy |||||||
3460 AspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeu 479
Db |||||||
3585 CGGCTGCTTTTATGACGCTGGGCTCAGTCTATGCTTTGCGAGCCAGAAATTGGCTA 3644
Qy |||||||
3480 ProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSer 499
Db |||||||
3645 CCTGGCCCTGCTACCGGCACAGAGACTTTCAAGACTGCTTAACGACCAACCAACCACT 3704
Qy |||||||
3500 AenPheThrTrpThrGlyAlaSerLysTyrAsnLeuAenGlyArgGluSerIleLeu 519
Db |||||||
3705 AACTTTCTTGACAGCGCCAGCAAAATATATCTCAATGGCCGCGATCGCTGGTGAAT 3764
Qy |||||||
3520 ProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGly 539
Db |||||||
3765 CCAGGACCACTATGGCCAGTCACAAGGACGATGAAGAAAATTTTCCCTATGCACGCG 3824
Qy |||||||
3540 ValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMet 559
Db |||||||
3825 AATCAATATTTGGCAAGAGGACCAACCGCAAGTAACGCGAGAATTAGATAATGTAATG 3884
Qy |||||||
3560 IleThrAspGluGluGluIleLysAlaThrAenProValAlaThrGluArgPheGlyThr 579
Db |||||||
3885 ATTACGGATGAAGAGAGATTCTGACCAATCTCTGTGGCAACAGACGATATGGAAT 3944
Qy |||||||
3580 ValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMet 599
Db |||||||
3945 GTGGCAATAAATTCAGAGCTCAAAATACAGCTCCACGACTAGAACTGTCAATGATCAG 4004
Qy |||||||
3600 GlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrp 619
Db |||||||
4005 GGGGCTCTACCTGGCATGGTGTGGCAAGATCGTGCATCTCTCTCTGATGGAGGCTTGG 4064
Qy |||||||
3620 AlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGly 639
Db |||||||
4065 GCAAGAATTCTCTCACAGGATGGACACTTTTCACTCTCTCTCTGATGGAGGCTTGG 4124
Qy |||||||
3640 LeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsnProPro 659
Db |||||||
4125 CTGAACACTCCGCTCTCTCAATCATGATCAAAATATCTCCGGTACCGGCAATCCTCG 4184
Qy |||||||
3660 AlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnVal 679

```

```

Db 4185 ACAGCTTTTACCGCCGCGCAAGTTTGTCTTATCTATCTCAGTACTCCACTGGACAGTC 4244
Qy 680 SerValGluIleGluTyrGluLeuGlnLysGluAsnSerLysArgTyrAsnProGluVal 699
Db 4245 AGCGTGAATAATTGATGGGAGCTACAGAAAGAAACAGCAACGTTGGAATCCAGAT 4304
Qy 700 GlnTyrThrSerAntyrrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGly 719
Db 4305 CAGTACACTTCCAACTACAACAGTCTGTTAATGTGACCTTTACTGTAGACACTAATGGT 4364
Qy 720 LeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 4365 GTTTATATGAACCTCGCCCTATTGGAACCGGCTATCTCACAGAAACTTG 4415

RESULT 4
LOCUS AVU48704
DEFINITION Adeno-associated virus 3 nonstructural protein and capsid protein
genes, complete cds, and complete genome.
ACCESSION U48704
VERSION U48704.1 GI:1408467
KEYWORDS Adeno-associated virus 3
SOURCE Adeno-associated virus 3
ORGANISM Adeno-associated virus 3
VIRUSES; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
REFERENCE 1 (bases 1 to 4726)
AUTHORS Muramatsu, S., Mizukami, H., Young, N.S. and Brown, K.E.
TITLE Nucleotide sequencing and generation of an infectious clone of
adeno-associated virus 3
JOURNAL Virology 221 (1), 208-217 (1996)
MEDLINE 96266430
PUBMED 8661429
REFERENCE 2 (bases 1 to 4726)
AUTHORS Muramatsu, S. and Brown, K.E.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-1996) Shin-ichi Muramatsu, Hematology Branch,
NHLBI/NIH, 9000 Rockville, MD 20892, USA
FEATURES
Location/Qualifiers
1..4726
/organism="Adeno-associated virus 3"
/mol_type="genomic DNA"
/strain="3H"
/db_xref="taxon:46350"
318..2192
/codon_start=1
/product="nonstructural protein"
/protein_id="AAC55048.1"
/db_xref="GI:1408468"
/translaton="MPGFYEIVLKVPSDLDERLPGLINSFVNWVAEKEDVPPDSMD
PNIEQAPLVAEKLQREFLVWRVRSKAPALFFVQFEGETYFHLHLVLIETIGVKS
MVGRVYVQIKELVTRIRYGVPEQLPNFPAVTKTRNGAGGKGVVDDCIPTNLLPK
TPELOWATNMDOYLSACLAEKRLVAQHLTHVQTOENKQENQNPNSDAPVRS
KTSARVMEYVGLVDGRTSEKOWIQEDQASYISFNAASRSOIKAALDNASKMSL
TKTAPDYLGSNPEDITKRIYQIELANGYDQYAAASVFLGWAQKKGKENTWLEK
PATTGKTNIAEAAIAHAFVFFGVNTNENFPFNDVDMVNWEEGKMTAKVVSACA
ILGSKRVDPCKSSAQIEPTVITSNTMCAVIDGNSITTFHQQLQDRMEPEL
TRLIDHDFKVTQEDKDFVWASDHVTVAHIEFYRKGAKKGPAPNDASPEKRE
CTSLAOPTTGDAAPADYADYQNKCSHRVGMMLPCKTCERMNQISNVCPTHGR
DCBCEPFGMSQSPVSVVKKTKYQLCPHILGRAPEIACACDLANVLDLDCVSEQ
"
2209..4419
/codon_start=1
/product="capsid protein"
/protein_id="AAC55049.1"
/db_xref="GI:1408469"
/translaton="MAADGYLPDWNLSGIREWALKGVQPPKANOQHDNRRG
LVLPYGLTGFNGLDKGEPEVNEDEAAALAEHDKAYDQKAGNPYLKYNHDAEFQE
RLQEDTFFGNGLRGAVQAKRILEPLGLVEAAATAPKKGKGVQDPQSPDSGSGV
KSGQPARQLNFGQDGSVPDQPLGEPAPATSLGNTWASGGAPNADNNEGA
DGVNSGNHCDQSLMDRVTITSTRWALPTYNHLYKQISSQSGASNDNHYFGS
TPWGYFDNFPHCFPSYDMORLNNNNNGFRPKLSFKLENIQVRGVTQNDQTTIAN
NLISTVOFTDSEYQLPYLGSAGHQLPPEPADVFMVPOXYGLTLANGSOAGRSSP
YCLTFPSQMLRTGNQFSTTFEDVFPHSSIAHSQSLDRMLNPLIDQLYLYLNRQSG

```


TTSGTNTQSRLLFFSQAGPQSMQLQARNWLPQCYRQORLSKTDNDNNNSNPPWTAASK
 YHLNDRSLVNPSPGAMSHKDDDEKFFPMHGNLIIPGEGTTASNAEILDNVMTIDEEEI
 RTTNPVAPQYGTGVANNLSNSTAPTCTGVNHOGLPGMWQDRDVLQGPIMAKIPIH
 TDGHPHSPGLMGFLKHPPIQIMKTPVPANPPTTSPAKFASFILTOYSTGVSVSE
 IEMELOKSNKRWPEIQYTSNYKSNVNDFTVDTNGVYSEPRPIGRILYLRNL"

BASE COUNT 1200 a 1295 c 1251 g 980 t

ORIGIN

Alignment Scores:

Pred. No.: 2,19e-209 Length: 4726
 Score: 3494.00 Matches: 637
 Percent Similarity: 91.72% Conservatives: 39
 Best local Similarity: 86.43% Mismatches: 59
 Query Match: 87.59% Indels: 2
 DB: 14 Gaps: 2

US-09-807-802a-13 (1-736) x AVU48704 (1-4726)

Qy 1 MetAlaIaAspGlyTyrLeuProAspTrpLeuGluAspAsnLeuSerGluGlyIleArg 20
 Db 2209 ATGGCTGTGCTACGGTATCTCCAGATTGGCTCGAGGACAACTTCTCGAAGGCATTCGT 2268
 Qy 21 GluTrpTrpAspLeuAspProGlyAlaProLysProLysAlaAsnGlnLysGlnAsp 40
 Db 2269 GAGTGGTGGCTCTGAACCTCGAGTCTCTCAACCCAAAGCGAACAACACACCGAGAC 2328
 Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
 Db 2329 AACCGTCCGGGTCTTGCTTCGGGTACAAATACCTCGACCCGGTAAACGACTCGAC 2388
 Qy 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
 Db 2389 AAAGAGAGCCCGTCAACGAGCGGACCGCGGCGCTCGAACACGACAAAGCTTACGAC 2448
 Qy 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
 Db 2449 CAGCAGCTCAGCGCGGTGACAAACCGTACCTCAAGTACACACCGCGACCGCGAGTTT 2508
 Qy 101 GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln 120
 Db 2509 CAGGAGCGTCTTCAAGAGTACGCTCTTTTGGGGGCAACCTTTGGCAGAGCAGCTCTCCAG 2568
 Qy 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaPro 140
 Db 2569 GCCAAAAGAGGATCTTGAGCTCTTGCTGTGGTTGAGGAGCAGCTAAACCGCTCT 2628
 Qy 141 GlyLysLysArgProValGlnGlnSerProGlnGluProAspSerSerSerGlyIleGly 160
 Db 2629 GGAAAGAGGGGCTGTAGTCACTCTCTCAGGAACCGGACTCATCATCTGCTGTGGC 2688
 Qy 161 LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
 Db 2689 AAATCGGGCAACAGCCTGCCAGAAAAGACTAAATTTTCGTACAGCTGGAGACTCAGAG 2748
 Qy 181 SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaValGlyPro 200
 Db 2749 TCAGTCCCAAGACCTCAACCTCTCGAGAACACCAACAGCAGCCGCCCAAGTTTGGGATCT 2808
 Qy 201 ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGly 220
 Db 2809 AATACATAGCTTTCAGCGGTGGCGCACCAATGGCAGACAAATACAGGGGTCCCATGGA 2868
 Qy 221 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIle 240
 Db 2869 GTGGGTAAATCTCAGGAATTTGGCATTCGGATTCCCAATGCTGGCGACAGAGTCAATC 2928
 Qy 241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIle 260
 Db 2929 ACCACAGACCAACCACTTGGGCCCTCGCCACTTACCAACACCACTCTCTACAGCAATC 2988
 Qy 261 SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrp 280
 Db 2989 TCCAGCCAAATCA---GGAGCTTCAAACGACAAACCACTACTTTGGCTACAGCACCCCTTGG 3045

Qy 281 GlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu 300
 Db 3046 GGGTATTTTACATTAAACAGATTCCACTGCCACTTCTCACACCGTACTGCGCAGCGACTC 3105
 Qy 301 IleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGln 320
 Db 3106 ATTAACAACAACCTGGGATTCGGGCCCAAGAAACTCAGCTTCAAGCTCTTCAACATCCNA 3165
 Qy 321 ValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThr 340
 Db 3166 GTTAGAGGGTTCACGACAGACGATGGCAGCAGCTATTGCAATTAACCTTACACGACAG 3225
 Qy 341 ValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGln 360
 Db 3226 GTTCAAGTGTTCAGGACTCGGAGTATCAGCTCCGCTACGTGCTCGGCTCGGCGCACCA 3285
 Qy 361 GlyCysLeuProPheProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr 380
 Db 3286 GGCTGTCTCCCGCGCTTTCAGCGGAGCTCTTCATAGTTCCTCAGTATGGATATCCTCACC 3345
 Qy 381 LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 400
 Db 3346 CTGAACAAACGGAAGTCAAGCGGTGGGACGCTCATCTCTTACTGCTGCGAGTACTTCCT 3405
 Qy 401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 420
 Db 3406 TCGCAGATGCTAAGGACTGGAAATAACTTCCAAATCAGCTATACCTTCGAGGATGTACCT 3465
 Qy 421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
 Db 3466 TTTCAACAGCAGCTACGCTCACAGCCAGAGTTTGGATCGCTTGGTAAATCTCTTATTAT 3525
 Qy 441 GlnTyrLeuTyrTyrLeuAsnArgThrGln---AsnGlnSerGlySerAlaGlnAsnLys 459
 Db 3526 CAGTATCTGTACTACTTGAACAGACGCGAAGAACACCTCTCGNACACCAACCAATCA 3585
 Qy 460 AspLeuLeuPheSerArgLysProAlaGlyMetSerValGlnProLysAsnTrpLeu 479
 Db 3586 CGGCTGTCTTTTAGCGAGCTGGGCTCAGTCTATGCTTTCAGCGCCGAGAAATTTGGTGA 3645
 Qy 480 ProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSer 499
 Db 3646 CCTGGGCTCTCTACCGGCAACAGAGACTTTCAAAGACTGCTAACGACGACCAACAACT 3705
 Qy 500 AsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsn 519
 Db 3706 AACTTCTTGGACAGCGGCGCAGCAATATCATCTCAATGGCGGAGCTCGCTGGTGAAT 3765
 Qy 520 ProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGly 539
 Db 3766 CCAGGACCGACTATGGCCAGTCAACAGGACGATGAAGAAAAATTTTCCCTATGACCGGC 3825
 Qy 540 ValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAsnValMet 559
 Db 3826 AATCTAATATTTGGCAAGAAAGGAGGACACGCGCAGTAAACGAGAAATAGATAATGATG 3885
 Qy 560 IleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThr 579
 Db 3886 ATTACGATGAAGAAGAGATTTCGTACCACCAATCTCTGTGGCAACAGACAGCAGTATGAA 3945
 Qy 580 ValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMet 599
 Db 3946 GTGGCAATAAATCTGACAGAGCTCAATACAGCTCCACAGCTGGAACCTGTCAATCATCAG 4005
 Qy 600 GlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrp 619
 Db 4006 GGGGCTTACCTGGCATGGTGGCAAGATCGTGACGTGTACCTTCAAGGACCTATCTGG 4065
 Qy 620 AlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGly 639
 Db 4066 GCNAAGATTCCTCACCGGATGGACACTTTTCATCTCTCTCTCTCTGATGGAGGCTTTGGA 4125

```

QY      640 LeuLysAsnProProGlnIleLeuLysAsnThrProValProAlaAsnProPro 659
DB      4126 CTGAACATCCGCCCTCTCAATCATGATCAAAATATCTCCGGTACCGCAATCCTCCG 4185

QY      660 AlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlnVal 679
DB      4186 ACAGCTTTCAGCCCGCCCAAGTTTGTCTCATTTATCATCTCACTACTCCACGAGTC 4245

QY      680 SerValGluLeuGluTyrProGluLeuGlnLysGluAsnSerLysArgTyrAsnProGluVal 699
DB      4246 AGCGTGGAAATTCAGTGGAGCTACAGAAAGAAACAGCAACGTTGGAATCCAGAGAT 4305

QY      700 GlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnGlnGly 719
DB      4306 CAGTACACTTCAACTACAAACAGTCTGTTATGTGGACTTTACTGTAGACACTAATGTT 4365

QY      720 LeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
DB      4366 GTTTATAGTGAACCTCGCCCTATTGGAACCCGCTATCTCACAGAACTTG 4416

RESULT 5
LOCUS   AY242997
DEFINITION Non-human primate Adeno-associated virus isolate AAVrh.8 capsid
ACCESSION AY242997
VERSION   AY242997.1 GI:29650489
KEYWORDS
SOURCE   Non-human primate Adeno-associated virus
ORGANISM Non-human primate Adeno-associated virus
REFERENCE 1 (bases 1 to 2211)
AUTHORS   Gao, G., Alvira, M.R., Somanathan, S., Lu, Y., Vandenbergh, L.H.,
           Rux, J.J., Calcedo, R., Sanmiguell, J., Abbas, Z. and Wilson, J.M.
TITLE     Adeno-associated viruses undergo substantial evolution in primates
           during natural infections
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 100 (10), 6081-6086 (2003)
PUBMED   12716974
REFERENCE 2 (bases 1 to 2211)
AUTHORS   Gao, G., Alvira, M.R., Somanathan, S., Lu, Y., Vandenbergh, L.H.,
           Rux, J.J., Calcedo, R., Sanmiguell, J., Abbas, Z. and Wilson, J.M.
TITLE     Direct Submission
JOURNAL   Submitted (23-FEB-2003) Department of Medicine, University of
           Pennsylvania School of Medicine and Wistar Institute, 204 Wistar
           Institute, 3601 Spruce Street, Philadelphia, PA 19104, USA
FEATURES
source   Location/Qualifiers
           1..2211
           /organism="Non-human primate Adeno-associated virus"
           /proviral
           /mol_type="genomic DNA"
           /isolate="AAVrh.8"
           /db_xref="taxon:226582"
           1..2211
           /gene="vpi"
           1..2211
           /gene="vpi"
           /codon_start=1
           /product="capsid protein"
           /protein_id="AA088183.1"
           /db_xref="GI:29650490"
           /translations="MAADGYLPDLENLSEGIREWMDLKPAPKPAKQKQDQGRG
           LVLPQVYLGPFNGLDKEPVNAADAAAEHDKAYDQOLKAGDNPVLYRNHADAEFOE
           RLEDRTSGFNLGRAVPQAKRVLPLGLVEEGAKTAPKRPVEQSPQEPDSSSGIG
           KTGQCPAKKLNFGTQDSSEVPDQPLGEPFPAAPSGLGENTVWAGGAFMADNNGA
           DVGNSGNHCDSTWLDSDVITTSRTWALPTYNHLYKQISNGTSGSTNDNTYFG
           YSTPMGYFDNRHFCHFSPPDQWRLLNNWGFPRKLNFLFNIOVKETVNEGTXTI
           ANFLTQVPTDSEYOLPYVLGSAHOCPLFPFPADVFPVQYILTLLANGSQAIGRS
           SFYCLEYFSPOMLRTGNNFQFSYTFEEDVFPFSSVYHSSQSLDRMLNPLIDQLYLVRT
           QTTGTGTQTLASQAGFSSKANQARNWVPCYRQORVSTTTNNQNNNFNFWTGAAK
           FKLNGRDLSLNPVGAASHKDDDDRRFPSSGVLIIFGQAGNDGVDIYSQVLIIDEEI
           KATNPVATEYGAIVAINNQAATQAGTGLVHNGQVTFPGWVWQNRDVLGGPIWAKLIPH
           TDGNFHPSPLMGGFGFKHPFQILLIKNTVPADPPLTFNQAOKLNSITQYSTQGVSE
           IEWELQKSNRWNPETQYTSNYLYKSTNVDFAVNTGCVSEPRPIGTRVLTNL"

```

```

BASE COUNT      573 a      668 c      572 g      398 t
ORIGIN
Alignment Scores:
Pred. No.:      1.7e-208      Length:      2211
Score:          3473.00      Matches:      631
Percent Similarity: 91.59%      Conservative: 44
Best Local Similarity: 85.62%      Mismatches: 60
Query Match:    87.06%      Indels: 2
DB:             14      Gaps: 2
US-09-807-802A-13 (1-736) x AY242997 (1-2211)
QY      1 MetAlaAlaAspGlyTyrLeuProAspTyrLeuGluAspAsnLeuSerGluGlyIleArg 20
DB      1 ATGCTGCCGATGGTTATCTTCAGATTGGCTCGAGNCAACCTCTCTGAGGGCATTCGC 60

QY      21 GluTyrTrpAspLeuLysProGlyAlaProLysProLysAlaAsnGlnGlnLysGlnAsp 40
DB      61 CAGTGGTGGGACTTGAAACCTGGAGCCCGAAACCCAAAGCCAAACCCAGCAAAAGCAGGAC 120

QY      41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
DB      121 GACGCGCGGGTCTGGTCTCTCTGGCTTACAGTACCTCGGACCCCTTCAACGGACTCGAC 180

QY      61 LysGlyGluProValAsnAlaAlaAspAlaAlaLalaLeuGluHisAspLysAlaTyrAsp 80
DB      181 AAGGGGAGGCCGCTCAACGCGCGGACGACGCGCCCTCGAGCAGCAGCAAGGCTACGAC 240

QY      81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
DB      241 CAGCAGCTCAAAAGCGGGTGACATCCGTACCTGGGTATATACACGCCGCGCGGAGTTT 300

QY      101 GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln 120
DB      301 CAGGAGCGCTGCAAGAAGATACGCTCTTTGGGGCAACCTCGGGCGGAGCAGTCTTTCAG 360

QY      121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGlyAlaLysThrAlaPro 140
DB      361 GCCAAGAGCGGGTCTCGAACCTCTCGGTCTGGTTGAGGAGGCGGCTAAGACGCTCCT 420

QY      141 GlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSerGlyIleGly 160
DB      421 GGAAGAAGAGAGACCGGTAGAGCAGTCGCGCACAAAGAGCCAGACTCTCTCTCGGSCATCGGC 480

QY      161 LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
DB      481 AAGACAGGCCAGCAGCCCGCTAAAAGAGACTCAATTTTGGTCAGACTGGCAGCTCAGAG 540

QY      181 SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro 200
DB      541 TCAGTCCCGCAGCCACACCTCTCGAGAACCTTCAGAGAACCTTCAGGAGCCCTCAGGTCTGGGACCT 600

QY      201 ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGly 220
DB      601 AATACAATGGCTTCAGCGCGGTGGCGCTCCAATGGCAGACAATAACGAAGCGCCGACGGA 660

QY      221 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIle 240
DB      661 GTGGGTAAATTCCTCGGGAAATTTGGCATTTGCATTTCCATGCTGGGGGACAGAGTCATC 720

QY      241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIle 260
DB      721 ACCACGAGCAGCCGGAACCTGGGCGCTCCACCTACCAACCAACCCCTCTACAAGCAATC 780

QY      261 SerSerAlaSerThrGlyAlaSer---AsnAspAsnHisTyrPheGlyTyrSerThrPro 279
DB      781 TCACACGCGCAGCTCGGGAGGAAGACCAACGACCAACACCTATTTTGGCTACAGACCCCC 840

QY      280 TrrGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrrGlnArg 299
DB      841 TGGGGGTATTTTGTGACTTCAACAGATTCCACTGTCTCTTTTACCACGCTGACTGGCAACGA 900

```

```
Qy 300 LeuileAenAenThrGlyPheAargProlysaArgLeuAenPheLyLeuPheAenile 319
Db 901 CTCAATCAACAAATTCGGGATTCGGCCCAAAAGACTCAACTTCAAGCTGTTCAACATC 960
Qy 320 GlnVallyGluValThrThrAenAenGlyValThrThrileAlaAenAenLeuThrSer 339
Db 961 CAGGTCAAGGAGTTCACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1020
Qy 340 ThrValGlnValPheSerAsePserGlyThrGlnLeuProTyValLeuGlySerAlaHis 359
Db 1021 ACCGTGCGAGTCTTTACGAGCTCGGAGTACAGTTACGTTACGTTACGTTACGTTAC 1080
Qy 360 GlnGlyCysLeuProProPheProAlaAsePvalPheMetileProGlnInTyGlyLeu 379
Db 1081 CAGGGATGTCCTCCCTCCGCGGACGCTTCATGCTTCATGCTTCATGCTTCATGCTTC 1140
Qy 380 ThrLeuAenAenGlySerGlnAlaValGlyArgSerSerPheTyCysLeuGluTyPhe 399
Db 1141 ACTTTAAACAATGAGACCAAGCCCTGGAGCTTCCTCTCTACTGCTGAGGATATTC 1200
Qy 400 ProSerGlnMetLeuAargThrGlyAenAenPheThrPheSerTyThrPheGluGluVal 419
Db 1201 CCATCGCAGATGCTGAGAACCGGCAACAACCTTTCAGTTACGTTACCTTCGAGCAGTG 1260
Qy 420 ProPheHisSerSerTyAlaHisSerGlnSerLeuAsePargLeuMetAenProLeulle 439
Db 1261 CCTTTTCACAGCAGCTACGCGCACAGCGACGCTGACAGGCTGATGAATCCCTCATC 1320
Qy 440 AspGlnTyLeuTyTyTyLeuAenAargThrGlnAenGlnSerGlySerAlaGlnAenLys 459
Db 1321 GACCACTACCTGTTACTCTGTTGTCAGAACCGCAACG---ACTGGAACCTGGAGGACCGAG 1377
Qy 460 AspLeuLeuPheSerAargGlySerProAlaGlyMetSerValGlnProLysAenThrLeu 479
Db 1378 ACTCTGCATTCAGCCAAAGCGGCTCTAGCTCAATGCGCAACACGAGGTAGAAATGGGTG 1437
Qy 480 ProGlyProCysTyAargGlnGlnAargValSerLysThrLysThrAseAenAenSer 499
Db 1438 CCGGAGCTTGTCTACCGCGCAGCAGCGCTCTCCACGACACCAACCAAGAACCAACAGC 1497
Qy 500 AenPheThrTrpThrGlyAlaSerLysTyAenLeuAenGlyAargGlySerIleAen 519
Db 1498 AACTTTGCTTGACGGGAGCTGCCAAGTTTAAGCTGAACGGCCGAGACTCTCTAATGAAT 1557
Qy 520 ProGlyThrAlaMetAlaSerHisLysAsePargGlyAsePhePheProMetSerGly 539
Db 1558 CCGGCGCTGGCAATGGCTTCCCAAGATGACGACGACCGCTTCTTCCCTTCGAGCGGG 1617
Qy 540 ValMetilePheGlyLysGlySerAlaGlyAlaSerAenThrAlaLeuAsePvalMet 559
Db 1618 GTCTGATTTTGGCAAGCAGGAGCGCGGGAACGATGAGTGGATTCAGCCAAAGTCTGT 1677
Qy 560 IleThrAsePgluGluGluLeuLysAlaThrAenProValAlaThrGluAargPheGlyThr 579
Db 1678 ATTACAGTGAAGGAATACTAAGCTACCAACCCCTGGCCACCAAGAAATATGAGCA 1737
Qy 580 ValAlaValAenPheGlnSerSerSerThrAsePProAlaThrGlyAsePvalHisAlaMet 599
Db 1738 GTGGCCATCAACCAACGAGCGCCCAATACGACGAGCGGACGACCGGACTCGTGCAACAC 1797
Qy 600 GlyAlaLeuProGlyMetValTrpGlnAsePargAsePvalTyLeuGlnGlyProileTrp 619
Db 1798 GGGGTGATTTCCCGGATGGTGTGGCAGATAGACGCTGACCTGCGAGGGTCCCATCTGG 1857
Qy 620 AlaLysileProHisThrAsePgluHisPheHisProSerProLeuMetGlyGlyPheGly 639
Db 1858 GCCAAATTCCTCACAGGACGCACTTTCACCGCTCTCCCTGATGGGCGGCTTTGGA 1917
Qy 640 LeuLysAenProProProGlnIleLeulleLysAenThrProValProAlaAsePProPro 659
Db 1918 CTGAAGCACCCTCTCTCAAAATTCATCAAGAACACACCGGTTCCAGCGGACCCGCGC 1977
Qy 660 AlaGluPheSerAlaThrLysPheAsePheileThrGlnInTySerThrGlyGlnVal 679
```

```
Db 1978 CTTACCTTCAACAGCAGCCAGCTGAATCTTTTCATCAGCAGTACAGCACCGAGGTC 2037
Qy 680 SerValGluIleGluTrpGluLeuGlnLysGluAenSerLysArgTrpAenProGluVal 699
Db 2038 AGGTGGAATCGAGTGGGAGCTGCGAAGAAAGAAACAGCAACGCTGGATCCAGATT 2097
Qy 700 GlnTyThrSerAenTyAlaLysSerAlaAsePvalAsePheThrValAsePvalAseGly 719
Db 2098 CAATACACTTCCAACTACTACAAATCTACAAATGTGAGCTTTGCTGTCAACAGGAGGG 2157
Qy 720 LeuTyThrGluProAargProIleGlyThrArgTyLeuThrArgProLeu 736
Db 2158 GTTATAGGAGGCTCGCCCATTTGCGACCCGTTACTTACCCTGCAACCTG 2208

RESULT 6
LOCUS AY243015 2217 bp DNA linear VRL 14-MAY-2003
DEFINITION Non-human primate Adeno-associated virus isolate AAVrh.10 capsid
protein (VPI) gene, complete cds.
ACCESSION AY243015
VERSION 1
KEYWORDS AY243015.1 GI:29650525
SOURCE Non-human primate Adeno-associated virus
ORGANISM Non-human primate Adeno-associated virus
REFERENCE 1 (bases 1 to 2217)
AUTHORS Gao,G., Alvira,M.R., Somanathan,S., Lu,Y., Vandenbergh,L.H.,
Rux,J.J., Calcedo,R., Sammiquel,J., Abbas,Z. and Wilson,J.M.
TITLE Adeno-associated viruses undergo substantial evolution in primates
during natural infections
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (10), 6081-6086 (2003)
PUBMED 12716974
REFERENCE 2 (bases 1 to 2217)
AUTHORS Gao,G., Alvira,M.R., Somanathan,S., Lu,Y., Vandenbergh,L.H.,
Rux,J., Calcedo,R., Sammiquel,J., Abbas,Z. and Wilson,J.M.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-2003) Department of Medicine, University of
Pennsylvania School of Medicine and Wistar Institute, 204 Wistar
Institute, 3601 Spruce Street, Philadelphia, PA 19104, USA
FEATURES
Location/Qualifiers
1..2217
/organism="Non-human primate Adeno-associated virus"
/proviral
/mol_type="genomic DNA"
/isolates="AAVrh.10"
/db_xref="taxon:226582"
1..2217
/gene="VPI"
1..2217
/gene="VPI"
/codon_start=1
/product="capsid protein"
/protein_id="AAO88201.1"
/db_xref="GI:29650526"
/translation="MAADGYLPDWLENDLSEGREWDLKPGAPKPAKXKQKDDGRG
LVLPQYLGPNGLDKEPVNAADAALEHDKAYDQKAGDNPIRLYNHDAEFQE
RLQEDTSFVGLRAVFAKRLVLEGLVEEGAKTAPGKRVPSPQSDSTGI
GKQGPQAKHRLNFGQTSVDPDPIGEPGAGSGTMAAGGAPMADNNEG
ADGVSSGNHNCDSITLDRVITSTRTWALPTYNHLYKQISNGTSGSTNDNTYF
GYSTPWGYPDNRFCHFSPRDMORLINNNMGPRKRLNFKLPIOVKEVTONEGKT
IANNLTSTQYVTDSEYQIYVLGSAHQSLPPFPADPMIPIGYLTLNNGSQVGR
SFFCYUEYQVQLRGNPFESYQEDVFFHSSYAHSLQDLKMLPLIDQYLYLSR
TSTGTGATGQALLFSQAGPNNSAQNKLPGCYRQQRVSTLTSNNNSNFWTGA
TKYHLNDRSLNVPVADNLOQNAAPIVGAVNSQALPGMWNQNDVYSLVLTSEE
EIKTNPVATEQYGVVADNLOQNAAPIVGAVNSQALPGMWNQNDVYSLVLTSEE
PHTDGNFPLMGGGLKHPPOILIKNTPPADPTTFSQAKLASFITQYSTGVGS
VELEWELQENSKRNPEIQSYNSYKSTNVDFAVNTDGTYSFPRIGTRYLRNL"
BASE COUNT 563 a 673 c 565 g 416 t
ORIGIN
Alignment Scores: 6.27e-207 Length: 2217
Pred. No.:
```

Score: 3448.00 Matches: 627
 Percent Similarity: 91.19% Conservative: 46
 Best Local Similarity: 84.96% Mismatches: 63
 Query Match: 86.44% Indels: 2
 DB: 14 Gaps: 2

US-09-807-802A-13 (1-736) x AY243015 (1-2217)

Qy 1 MetAlaAlaAaspGlyTyrLeuProAspThrLeuGluAaspAenLeuSerGluGlyIleArg 20
 Db 1 ATGGCTGGCGATGGTTATCTCCAGATTGGCTCGAGGACAACCTCTCTGAGGGGCAATTCGC 60

Qy 21 GluTptTfPaspLeuLeuProGlyAlaProLysProLysAlaAasnGlnGlnLysGlnAap 40
 Db 61 GAGTGGTGGGACTTGAACTTGAGCCCGGCAACCCAAAGCCAAACAGCAAAAGCAGGAC 120

Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAasnGlyLeuAap 60
 Db 121 GACGGCCGGGGTCTGGTCTCTCGCTACAAAGTACCTCGGACCCCTTCAACGGACTCGAC 180

Qy 61 LysGlyGluProValAasnAlaAaspAlaAlaLeuGluHiAaspLysAlaTyrAap 80
 Db 181 AAGGGGGAGCCCGTCAACGGCGGACGCGGGCCCTCGAGCAGCAAGAGCCCTTACGAC 240

Qy 81 GlnGlnLeuLysAlaGlyAaspAenProTyrLeuArgTyrAasnHiAalAaspAlaGluPhe 100
 Db 241 CAGCAGCTCAAGCGGGTGACATCCGTACCTCGGGTATACCAACCGCGAGCCGAGTTT 300

Qy 101 GlnGluArgLeuGlnGluAapThrSerPheGlyGlyAasnLeuGlyArgAlaValPheGln 120
 Db 301 CAGGAGCGTCTCAAGAGATACGCTTTTGGGGCAACCTCGGGCGAGCAGTCTTCCAG 360

Qy 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaPro 140
 Db 361 GCCAAGAAGCGGGTCTCGAAGCTCTCGGTCTGGTTGAGGAAGGGCGCTAAGACGGCTCT 420

Qy 141 GlyLysLysArgProValGluGlnSerProGln---GluProAaspSerSerSerGlyIle 159
 Db 421 GGNAGAAGAGACCGGTAGAGCCATCACCCAGCGTTCTCCAGACTCTCTACGGGCATC 480

Qy 160 GlyLysThrGlyGlnGlnProAlaLysLysArgLeuAasnPheGlyGlnThrGlyAaspSer 179
 Db 481 GGCAGAAAGAGCCGAGCAGCCCGGAAAAAGAGACTCAACTTTGGGCGAGCTGGCGACTCA 540

Qy 180 GluSerValProAaspProGlnProLeuGlyGluProAlaThrProAlaAlaValGly 199
 Db 541 GAGTCAAGTCCCGACCCCTCAACCAATCGGAGAACCCCGCGAGCCCTCTCGTCTGGGA 600

Qy 200 ProThrThrMetAlaSerGlyGlyAlaProMetAlaAaspAasnGlnGlyAlaAap 219
 Db 601 TCTGGTACAAATGGCTCGAGCGGTGGCGCTCCAAATGGCAGACAATAACGAAGCGCCGAC 660

Qy 220 GlyValGlyAasnAlaSerGlyAenThrHisCysAaspSerThrTripleuGlyAaspArgVal 239
 Db 661 GGAGTGGGTAGTTCCTCAGGAAATGGGCATTTGGCATTTGCATATGGCTGGGGCGACAGATC 720

Qy 240 IleThrThrSerThrArgThrTptAlaLeuProThrTyrAasnHiAalLeuTyrLysGln 259
 Db 721 ATCACCACGAGCAGCCCGAACCTGGGGCCCTCCCGACCTACACCAACCACTCTACAGCAA 780

Qy 260 IleSerSerAlaSerThrGlyAlaSer---AasnAaspAasnHiTyrPheGlyTyrSerThr 278
 Db 781 ATCTCCAACGGGACTTCGGGAGGAGACCAACGACCAACACCTACTCTCGGCTACAGCAC 840

Qy 279 ProTptGlyTyrPheAaspPheAasnArgPheHisCysPheSerProArgAapThrGln 298
 Db 841 CCTGGGGGTATTTTACATTTTAAACATTTCCACTTCCACTCTCACACGCTGACTGGCAG 900

Qy 299 ArgLeuIleAasnAasnTptGlyPheArgProLysArgLeuAasnPheLysLeuPheAasn 318
 Db 901 CGACTCATCAACAAACAACTGGGGATTCGGGGCCCAAGAGACTCAACTTCAAGCTCTTCAAC 960

Qy 319 IleGlnValLysGluValThrAasnAaspGlyValThrThrIleAlaAasnAenLeuThr 338

Db 961 ATCCAGGTCAAGGAGGTCAAGCAGATGAAGGCCAACCAAGACCATCGCAATAACCTTACC 1020

Qy 339 SerThrValGlnValPheSerAaspSerGluTyrGlnLeuProTyrValLeuGlySerAla 358
 Db 1021 AGCACGATTTCAGGTCTTTTACGGACTCGGAATACACAGCTCCGCTCTCTCTCTCTGCG 1080

Qy 359 HisGlnGlyCysLeuProPheProAlaAaspValPheMetIleProGlnTyrGlyTyr 378
 Db 1081 CACCAGGGCTGCCTGCTCTCGTTCCCGCGGAGCGTCTTCATGATTCTCTCAGTACGGGTAC 1140

Qy 379 LeuThrLeuAasnAasnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyr 398
 Db 1141 CTGACTCTGAACAATGCGAGTCAGCGCGTGGCGGTCTCTCTCTCTCTCTCTCTCTCTGAG 1200

Qy 399 PheProSerGlnMetLeuArgThrGlyAasnAasnPheThrPheSerTyrThrPheGluGlu 418
 Db 1201 TTTCTCTTCTCAATGCTGAGAACGGGCAACACTTTTGAATTCAGTACCATCGATTGAGGAC 1260

Qy 419 ValProPheHisSerSerTyrAlaHisSerGlnSerLeuAaspArgLeuMetAasnProLeu 438
 Db 1261 GTGGCTTTTCACAGCAGCTACGCGCACAGCCAAAGCTTGGACCGGCTGATGAACCCCTC 1320

Qy 439 IleAapGlnTyrLeuTyrTyrLeuAasnArgThrGlnAasnGlnSerGlySerAlaGlnAasn 458
 Db 1321 ATCGACCAAGTACCTGTACTACTCTCTCGGACTCAGTCTCCAGCGGAGGTACCGCAGGA 1380

Qy 459 LysAaspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAasnTrp 478
 Db 1381 CAGCAGTTGCTTATTTTCTCAGCGCGGCTTAATAACATGTCCGCTCAGCGCCCAAACTGG 1440

Qy 479 LeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAasnAasnAasn 498
 Db 1441 CTACCCGGGCTCTGTACCGGCAGCAACCGCTCTCCAGCACACTGTCGCAAAATAACAAC 1500

Qy 499 SerAasnPheThrTptThrGlyAlaSerLysTyrAasnLeuAasnGlyArgGluSerIleIle 518
 Db 1501 AGCAACTTTGCTGGACCGGTGCCCAAGATATCTCTGAATGGCAGAGACTCTCTGTA 1560

Qy 519 AasnProGlyThrAlaMetAlaSerHisLysAaspAapGluAaspLysPhePheProMetSer 538
 Db 1561 AATCCGGTGTGCTGTATGGCAACCCACAGAGCAGCAGAGAGCGATTTTTTCCGTCAGC 1620

Qy 539 GlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAasnThrAlaLeuAaspAasnVal 558
 Db 1621 GGAGTCTTAATGTTTGGAAACAGGAGCTCGAAAAGACAACAGTACTATAGCAGCGTT 1680

Qy 559 MetIleThrAaspGluGluGluIleLysAlaThrAasnProValAlaThrGluArgPheGly 578
 Db 1681 ATGCTAACCAAGTGGAGAAATTAACCAACCAACCCAGTGGCCACAGACAGTACCGGC 1740

Qy 579 ThrValAlaValAasnPheGlnSerSerSerThrAaspProAlaThrGlyAaspValHisAla 598
 Db 1741 GTGTGGGCGGATTAACCTGCAACAGCAAAACCCCGCTCTCTATTTAGGGGCGCTCAACAGT 1800

Qy 599 MetGlyAlaLeuProGlyMetValTptGlnAaspArgAapValTyrLeuGlnGlyProIle 618
 Db 1801 CAAGGAGCCTTACCTGGCATGCTTGGCAGAACCGGAGCGTGTACTCTCAGGGTCTCTATC 1860

Qy 619 TrpAlaLysIleProHisThrAaspGlyHisPheHisProSerProLeuMetGlyLysPhe 638
 Db 1861 TGGGCCAAGATTCCTCACACGAGCGAAACTTTTCACTCTCCCTCGCGCTGATGGAGGCTT 1920

Qy 639 GlyLeuLysAasnProProGlnIleLeuIleLysAasnThrProValProAlaAasnPro 658
 Db 1921 GGACTGAACACCCGCTCTCTCAGATCTCTGATTGAAGATACACCTGTTCCCGCGATCT 1980

Qy 659 ProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGln 678
 Db 1981 CCAACTTACCTTCAAGCTCAAGCTGAAGTGGCGTCTGTTCTATCAGCGAGTACAGCAGCAG 2040

Qy 679 ValSerValGluIleGluTptGluLeuGlnLysGluAasnSerLysArgTptAasnProGlu 698

Db 2041 GTACGCTGGAAATTGAATGGAGCTGCAGAAAGAAACACGACCTGCAACCCAGAG 2100
Qy 699 ValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsn 718
Db 2101 ATTCAATACACTTCCAACTACTACAAATCTACAATGTGGACTTTGCTGTTTAAACAGAT 2160
Qy 719 GlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 2161 GGCACTTATTCTGAGCTCGGCCCATCGCACCGCTTACCTCACCGCTAATCTG 2214

RESULT 7
AF513851 4721 bp DNA linear VRL 05-SEP-2002
LOCUS Adeno-associated virus 7 nonstructural protein and capsid protein
DEFINITION genes, complete cds.
ACCESSION AF513851 GI:22652859
VERSION AF513851.1
KEYWORDS Adeno-associated virus 7
SOURCE Adeno-associated virus 7
ORGANISM Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
REFERENCE 1 (bases 1 to 4721)
AUTHORS Gao,G.P., Alvira,M.R., Wang,L., Calcedo,R., Johnston,J. and Wilson,J.M.
TITLE Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (18), 11854-11859 (2002)
PUBMED 12192090
REFERENCE 2 (bases 1 to 4721)
AUTHORS Alvira,M.R.
TITLE Direct Submission
JOURNAL Submitted (20-MAY-2002) Institute for Human Gene Therapy,
University of Pennsylvania, M6.40 Maloney Bldg, 36th & Spruce Sts,
Philadelphia, PA 19104, USA
FEATURES
source location/Qualifiers
1..4721 /organism="Adeno-associated virus 7"
/proviral
/mol_type="genomic DNA"
/specific_host="rhesus monkey"
/db_xref="taxon:202812"
334..2205 /note="similar to AAV2 Rep 78 protein"
/codon_start=1
/product="nonstructural protein"
/protein_id="AA03854.1"
/db_xref="GI:22652860"
/translation="MPGEYEVIVKPSDLDEHLPGISDSFVNNVAEKWEWELPPDSMD
LNLEQPLTVAELKQDLQVWRVSKAPEALFFVQFEKESYFHLHLVETTVGVS
MVLGRFLQIRKLVQTYRGVPTLPNMFATKTRGAGGKGVVDECIYNYLLPK
TOPELQWANTMEYVISACLNLAERKRLVAOHLTHVQTOQNKENLNPNSDAPVRS
KTSARYMBELVGLDRGITSEKQWIOEDQASYISFNAAASNSRSQIKAAONAGKIMAL
TKSAPDYLVGSLPADIKTNRIYRILELNGVDPAVAGSVFLGMAQKFGKENTLWLG
PATGKTNIADIAIAHAVPPGCVNWTNENPPNCDVMKVLWEEGKTKAVVSASA
ILGSKVRDQKCKSSAQIDPTPVITSTNNMCAVDIGNSTTFEHQPLQDRMPKFL
TRLEHDFGKVTQGEVFEFFWASDHDVTEVAHFYVRKGGASKRPAPDDADISEPKKA
CPSVADPSTDAEGAVDFADRYQNKCSRHAGMIQMLFPCKTCERMNQNFICFTHGV
RDCLPCPGVSGSPVVRKTKYRKLCAIHLHGRAPEIACSDLVNVDLDDCVSEQ"
222..4435 /note="similar to AAV2 VP1 protein"
/codon_start=1
/product="capsid protein"
/protein_id="AA03855.1"
/db_xref="GI:22652861"
/translation="MAADGYPDLWLEDNLSEGIREWMDLKPAPKPKANQKQDNGRG
LVLPQYKYLGPNGLDKEPVNADAAALEHDKAYDQQLKAGDNPYLRYNHDAEFQE
RLQEDTSFGNGLRAVFAQKRVLEPLGLVEEGAKTAPAKRPVEPQSPQSPSTGI
GKGGQAPARKLQGTQDGSVPDPLGEPAPASVSGVTPAAGGAPADNNNG
ADVGNAENHCHDSTWLDGVITTTSTWALPTYNHLYKQISSETAGSTNDNTYFG
YSTPMGYDFNRFCHFPDRWQRLNNWGFPRKCLFKNLQVKEVTNDGVTTI
ANNLTSTIQVSDSEYQLPYVLGSAHOCGLPPFPADVFMIPQYGLTLNNGSQSVGRS
SFYCLGYPFQMLRQTNPFERSYFEDVPFFHSHAHOSLDRLNPLIDQLYLYIART
QSNPGGTAGNLTQYGGPSTMAEQAKNWLPGFCFROQRVSKTLQDQNNNSNFAWTGA

TKYHLNGRSLNVPGVAMATHKDDRPFPSGVLIFGKGTATNKTTLNVLMTNBBE
IPTNPVATERYGIYSSNLQAAANTAAQTQVNNQGLPGWNNQRNDVILQGPYAKIP
HTDGNPHSPGLMGFGFLGKHPPOILLKNTFPANPPEVTFPAKFASFITQYSTQVSV
EIEWELQKNSKRWNPETQYTSNFEKQGTGVDPAVDSQGVYSEPRPIGTRYLTRLN"
BASE COUNT 1108 a 1405 c 1301 g 907 t
ORIGIN
Alignment Scores:
Pred. No.: 3 68e-206 Length: 4721
Score: 3442.50 Matches: 630
Percent Similarity: 92.02% Conservative: 50
Best Local Similarity: 85.25% Mismatches: 54
Query Match: 86.30% Indels: 5
DB: 14 Gaps: 4
US-09-807-802A-13 (1-736) x AF513851 (1-4721)
Qy 1 MetAlaAlaAspGlyTyrLeuProAspTyrLeuGluAspAsnLeuSerGluGlyIleArg 20
Db 2222 ATGGCTGCCGATGGTTATCTTCAGATTGGCTCGAGACAACTCTCTGAGGCATTTCG 2281
Qy 21 GluTyrTrpAspLeuLysProGlyAlaProLysProLysAlaAsnGlnGlnLysGlnAsp 40
Db 2282 GAGTGGTGGACCTGAACCTGGAGCCCGAAACCCCAAGCCCAACAGCAAAAGCAGGAC 2341
Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
Db 2342 AACGGCCGGGGTCTGGTGTCTTCTGTGCTACAGTACCTCGACCCCTTCAACGGACTCGAC 2401
Qy 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
Db 2402 AAGGGGAGCCCTCAACCGCGGAGCGACGAGCGCCCTCGAGCAGCAAGGGCTCAGCAG 2461
Qy 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
Db 2462 CAGCAGCTCAAGGGGTGACAAATCGTACTCGGTATTAACCAACCGCGGCGGAGTTT 2521
Qy 101 GlnGluArgLeuGlnGluAspThrSerPheGlyLysLeuGlyArgAlaValPheGln 120
Db 2522 CAGGAGCGTCTGCAAGAAGATACGTCATTTGGGGGCAACCTCGGGCGAGCAGTCTTCAG 2581
Qy 121 AlaLysLysArgValLeuGluProLeuGlyLysValGluGlyAlaLysThrAlaPro 140
Db 2582 GCCAAGAAGCGGGTTCTCGAACCTCTCGGTCTGGTGTAGGAAGCGCTAAGACGGCTCT 2641
Qy 141 GlyLysLysArgProValGluGlnSerProGln---GluProAspSerSerSerGlyIle 159
Db 2642 GCAAGAAGAGACCGGTAGAGCGGTACCTCAGCGTTCCCCCGACTCTCTCCACGGGCATC 2701
Qy 160 GlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSer 179
Db 2702 GGCAAGAAGAGCCAGCAGCGCGCGCAGAAAGAGACTCAATTTTCGTTCAGACTGGCGACTCA 2761
Qy 180 GluSerValProAspProGlnProLeuGlyGluProAlaThrProAlaAlaValGly 199
Db 2762 GAGTCAGTCCCGACCTTCAACCTCTCGAGAACCTTCAGAGCGCCCTCTAGTGTGGGA 2821
Qy 200 ProThrThrMetAlaSerGlyGlyValProMetAlaAspAsnAsnGluGlyAlaAsp 219
Db 2822 TCTGGTACGTGGCTGCAGCGGTGGCGCACCAATGCAGACAAATAACGAAGGTGCCGAC 2881
Qy 220 GlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgVal 239
Db 2882 GGAGTGGGTAATGCTCAGGAATTTGGCAATTCACATGGCTGGCGGCGACAGAGTC 2941
Qy 240 IleThrThrSerThrArgThrTrpAlaLeuProThrTrpAsnHisLeuTyrLysGln 259
Db 2942 ATTACCAACGACACCCGAACTGGGGCCCTCCCGACCTTACCAACCAACCACTCTCAAGCAA 3001
Qy 260 IleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrPro 279
Db 3002 ATCTCCAGTGAACACTGCAGGTAGTACCAACGACACCACTTCTTCGGCTACAGCACCCCC 3061

QY 280 TrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTyrGlnArg 299
Db 3062 TGGGGGTATTTGACTTTAAACAGATTCCACTGCCACTTCTCACCACGCTGACTGGCAGCGA 3121
QY 300 LeuIleAsnAsnAsnTyrGlyPheArgProLysArgLeuAsnAsnPheLeuLeuPheAsnIle 319
Db 3122 CTCATCAACAACAACCTGGGGATTCGGGCCCAAGAGCTGCGGTTCAAGCTCTTCAACATC 3181
QY 320 GlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSer 339
Db 3182 CAGGTCAGGAGGTCAACAGCAATACACGGCGTTACGACCATCGCTTAATAACCTTACCAGC 3241
QY 340 ThrValGlnValPheSerAspSerGlnTyrGlnLeuProTyrValLeuGlySerAlaHis 359
Db 3242 ACGATTCAAGGTATTCGAGCTCGGAATACCAAGCTGCCGTACGCTCTCGGCTTCGGCAC 3301
QY 360 GlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeu 379
Db 3302 CAGGCTGCCTCGCTTCGGCGGACGCTTTCATGATTCCTCAGTACGGCTACCTG 3361
QY 380 ThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhe 399
Db 3362 ACTCTCAACAATGGCAGTCAGTCTGTGGACGTTCTCTTCTACTGCTGGAGTACTTC 3421
QY 400 ProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluVal 419
Db 3422 CCCTCTCAGATCTGAGACGGGCAACAATTTGATTCAGTACAGCTTCGAGGACGTG 3481
QY 420 PropheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIle 439
Db 3482 CTTTCCACAGCAGCTAGCACACAGCAGCCTGGACCGCTGATGAATCCCTCATC 3541
QY 440 AspGlnTyrLeuTyrTyrLeuAsnArgThrGln--AsnGlnSerCysSerAlaGlnAsn 458
Db 3542 GACCACTTCTACTTACCTGCGCCAGAACACAGAGTAACCCAGGAGGCACACTGGCAAT 3601
QY 459 LysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTyr 478
Db 3602 CGGGAACTGCAGTTTACAGGGCGGGCTTCAACTATGGCCGACACAGCCAGATTTGG 3661
QY 479 LeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 498
Db 3662 TTACTCTGAGCTTGTCTCGGCAACAAGAGTCTCAAAACGCTGGATCAAAACAACAAC 3721
QY 499 SerAsnPheThrTyrThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIle 518
Db 3722 AGCAACTTTGCTTGGACTGGTGCACCAATATCACTGAACGGCAGAACTCGTTGGTT 3781
QY 519 AsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSer 538
Db 3782 AATCCGGCGTCCGCATGGCACTCACAGGACGACGAGGACCGCTTTTCCATCCAGC 3841
QY 539 GlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsn---ThrAlaLeuAspAsn 557
Db 3842 GGAGTCTCTGATTTTGGAAAA-----ACTGGAGCAACTAAACAACATACATTGGAAAT 3895
QY 558 ValMetIleThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 577
Db 3896 GTGTTAATCAAAATAGAGAAGAAATTCGTCTACTTAATCTCTGTAGCCAGGAAGAAATAC 3955
QY 578 GlyThrValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHis 597
Db 3956 GGGATAGTCAGCAGCAACTTACAAGCGGCTTAATCTGCAGCCACAGACACAAGTTGTCAAC 4015
QY 598 AlaMetGlyAlaLeuProGlyMetValTyrGlnAspArgAspValTyrLeuGlnGlyPro 617
Db 4016 AACACAGGAGCCTTACTTGGCATGGTCTGGCAGAACCGGACGCTGTACCTGCAGGGTCCC 4075
QY 618 IleTyrAlaLysIleProHisThrAspGlyHisPheHisPheProLeuMetGlyGly 637
Db 4076 ATCTGGGCCAAGATTCCTCACACGATGGCAACTTTCACCCGCTCTCTCTTTGATGGCGGC 4135
QY 638 PheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsn 657

Db 4136 TTGGACCTTAAACATCCGCTCTCAGATCTGATCAAGAACACTCCCGTCCGCTAAT 4195
QY 658 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 677
Db 4196 CCTCGGAGGTGTTTACTCTCTGCCAAGTTTGTCTGTTCTATCACACAGTACAGCACCGGA 4255
QY 678 GlnValSerValGluIleGluTyrGluLeuGlnLysGluAsnSerLysArgTyrAsnPro 697
Db 4256 CAAGTCAGCTGGAATTCGAGTGGAGCTCGAAGGAACACAGAACGCTGGAACCCG 4315
QY 698 GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsn 717
Db 4316 GAGATTCACTACACCTTCAACTTTGAAAGACAGACTGGTGTGGACTTTGCCGTTGACAGC 4375
QY 718 AsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 4376 CAGGCTGTTACTCTGAGCCTCGCCCTATTGGCACTCTGTACCTCACCCGTAATCTG 4432
RESULT 8
AR264580 8179 bp DNA linear PAT 10-APR-2003
LOCUS AR264580 Sequence 5 from patent US 6491907.
ACCESSION AR264580
VERSION AR264580.1 GI:29692850
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 8179)
AUTHORS Rabinowitz, J.E., Samulski, R.J. and Xiao, W.
TITLE Recombinant parvovirus vectors and method of making
JOURNAL Patent: US 6491907-A 5 10-DEC-2002;
FEATURES Location/Qualifiers
1..8179
/organism="unknown"
BASE COUNT 2082 a 2078 c 2027 g 1992 t
ORIGIN
Alignment Scores:
Pred. No.: 1.24e-205 Length: 8179
Score: 3439.00 Matches: 626
Percent Similarity: 90.50% Conservative: 41
Best Local Similarity: 84.94% Mismatches: 68
Query Match: 86.21% Indels: 2
Gaps: 2
US-09-807-802A-13 (1-736) x AR264580 (1-8179)
QY 1 MetAlaAlaAspGlyTyrLeuProAspTyrLeuGluAspAsnLeuSerGluGlyIleArg 20
Db 2133 ATGGCTGCGATGGTTATCTTCAGATTGGCTCCAGGACACTCTCTCTGAAGGAATAAGA 2192
QY 21 GluTyrTrpAspLeuLysProGlyAlaProLysProLysAlaAsnGlnGlnLysGlnAsp 40
Db 2193 CAGTGTGGAGCTCAAACTGGCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 2252
QY 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
Db 2253 GACAGAGGGGTCTGTGCTTCTGGGTACAGTACCTCGGACCTTCAACGGACTCCAC 2312
QY 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
Db 2313 AAGGAGAGCGGTCAACAGGAGAGAGCGCGCGCTCGAGACGACCAAGACCTACGAC 2372
QY 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
Db 2373 CGGAGCTCGACGGGAGACACACCGTACCTCAAGTACAAACCAACCAACCGCGCGAGTTT 2432
QY 101 GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln 120
Db 2433 CAGGAGCGCTTTAAGAAAGATACGTCTTTTGGGGCAACCTTCGGACGAGAGCTCTCCAG 2492

121 AlalysLysArgValLeuGluProLeuGlyLeuValGluGlyAlaLysThrAlaPro 140
141 GlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerGlyIleGly 160
161 LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
181 SerValProAspProGlnProLeuGlyGluProProAlaLysValGlyPro 200
201 ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnGluGlyAlaAspGly 220
221 ValGlyAsnAlaSerGlyAsnThrHisCysAspSerThrTrpLeuGlyAspArgValIle 240
241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIle 260
261 SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrp 280
281 GlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu 300
290 GGGTATTGGTCTCAACAGATTCCACTGCCATTTTCCACACGCTGCTGCCAAGACTC 3029
301 IleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGln 320
3030 ATCAACAACAACTGGGGATTCGGACCCCAAGAGACTCAACTTCAAGCTCTTTAACTTCAA 3089
321 ValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnLeuThrSerThr 340
3090 GTCAAGAGGTCACGAGATGACGGTACGACGAGATTTGCAATTAACCTTTACCGACGC 3149
341 ValGlnValPheSerAspSerGlyTyrGlnLeuProTyrValLeuGlySerAlaHisGln 360
3150 GTTCAGGTGTTTACTGACTCGAGTACCAGCTCCGCTACGTTCCGCTCGGCGCACCAA 3209
361 GlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr 380
3210 GGCTGTCTCCGCGCGTTTCCAGCGGACGCTCTTCATGGTCCCTCAGTATGGATACCTCACC 3269
381 LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 400
3270 CTGAACAACGGAAGTCAAGCGGTGGAGCGCTCATCTTTTACTGCTGGAGTACTTCCCT 3329
401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 420
3330 TCGCAGATGCTAAGGACTGGAAATAACTTCCATTCAGCTATACCTTCGAGGATGTACCT 3389
421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
3390 TTTCACAGCAGCTAGCTCACAGCCAGAGTTTGGATCGCTTGGATGAATCTCTTTATTGAT 3449
441 GlnTyrLeuTyrTyrLeuAsnArgThrGln--AsnGlnSerGlySerAlaGlnAsnLys 459
3450 CAGTATCTGTACTACTCTGAACAGAACGACGAGAAACACTCTGGAAACCAACCAATCA 3509
460 AspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeu 479
3510 CGGCTGCTTTTACGAGCGCTGGGCTCAGTCTATGCTTTTGCAGGCCAGAGAAATTTGGCTA 3569
480 ProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSer 499

3570 CTTGGCCCTGTCTACCGCAACAGAGACTTTCAAAGACTGCTAACGACACACACAGT 3629
500 AsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleAsn 519
3630 AACTTTCTTGGACCGCCAGCAATATCATCTCAATGGCCGACCTCGTGGTGAAT 3689
520 ProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGly 539
3690 CCAGGACCACTATGCCCAGTCACAGGACGATGAAGAAAAATTTTCCCTATGCACGGC 3749
540 ValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMet 559
3750 AATCTAATATTGGCAAGAGGCAACGCGCAAGTAACGCAAGATTAGATAATGTAATG 3809
560 IleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThr 579
3810 ATTACGGATGAGAGAGATTGTCACCAACATCTCTGTGGCAACAGCAGATGTGNACT 3869
580 ValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMet 599
3870 GTGGCAATAACTTGCAGAGCTCAATAACAGCTCCCGACTGGAACCTGTCAATCATCAG 3929
600 GlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrp 619
3930 GGGGCCCTTACCTGGCATGGTGGCAAGATCGTGACGTGTACCTTCAAGGACCTATCTGG 3989
620 AlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGly 639
3990 GCAGAGATTTCTCACAGGATGGACACTTTCATCTCTCTCTGATGGAGGCTTTGGA 4049
640 LeuLysAsnProProGlnIleLeuLysAsnThrProValProAlaAsnProPro 659
4050 CTGAACATCCGCTCTCTCAATCATGATCAAAATATCTCCGCTACCTGCGAATCTCTTCG 4109
660 AlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnVal 679
4110 ACCACCTTTCAGTGGCGCAAGTTTGTCTCTCATCACAGTACTCCCGGACGAGGTC 4169
680 SerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluVal 699
4170 AGCTGGAGATTCAGTGGGNGCTGCAGAGAGGAAACAGCAACGCTGGATCCGAAAT 4229
700 GlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGly 719
4230 CAGTACACTTCCAACTACAACAGTCTGTTAATCGTGGACTTACCGTGGATCTAATATGC 4289
720 LeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
4290 GTGTATTTCAGAGCTCGCCCATTTGGCAGATACCTGACTGTAATCTG 4340

RESULT 9
AY243007
LOCUS
DEFINITION
Non-human primate Adeno-associated virus isolate AAVrh.2 capsid
protein (VPI) gene, complete cds.
ACCESSION
AY243007
VERSION
AY243007.1 GI:29650509
KEYWORDS
Non-human primate Adeno-associated virus
ORGANISM
Non-human primate Adeno-associated virus
VIRUSES; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
REFERENCE
1 (bases 1 to 2217)
Rux, J., Calcedo, R., Samiuel, J., Abbas, Z., and Wilson, J.M.
Adeno-associated viruses undergo substantial evolution in primates
during natural infections
Proc. Natl. Acad. Sci. U.S.A. 100 (10), 6081-6086 (2003)
12716974
2 (bases 1 to 2217)
Gao, G., Alvira, M.R., Somanathan, S., Lu, Y., Vandenberghe, L.H.,
Rux, J., Calcedo, R., Samiuel, J., Abbas, Z., and Wilson, J.M.
Direct Submission
TITLE

JOURNAL Submitted (23-FEB-2003) Department of Medicine, University of Pennsylvania School of Medicine and Wistar Institute, 204 Wistar Institute, 3601 Spruce Street, Philadelphia, PA 19104, USA

FEATURES
source Location/Qualifiers
1..2217
/organism="Non-human primate Adeno-associated virus"
/proviral
/mol_type="genomic DNA"
/isolate="AAVrh.2"
/db_xref="taxon:226582"
1..2217
/gene="VP1"
1..2217
/gene="VP1"
/codon_start=1
/product="capsid protein"
/protein_id="AA088193.1"
/db_xref="GI:29650510"
/translation="NAADGYLPDWLEDSLEIREWMDLPKPKPKANQKQDDGRG
LVLPGYKVLGPNGLDGEVYNADAAAEHDKAYDQOLKAGDNPYLRYNADAEFOE
RLQDTPSGNLGRAVFOAKKXVLEPLGLVEEGAKTA PKKRPVEPSPQSDSTGI
GKKHQAPARKLNFCQTGDSSEVPDPQIGPPAGPSGLSGTWAAGGAPADNNEG
ADGVSSGNHCDSTWLGDRTVITSTWALPTNNHLYKQISNGTSGGSDNTNYF
GYSTPWGTFDFRHFSPDWORLNNWFRKRLNFKLNIQVKEVTONSGTKT
IANNLTSTIQVDTSEYQLPYVPSGAHQCLPPFPADVPMIPOYGYLTNNQSOAVGR
SSFYCLEYPPSOMLRTGNPFSTYFEDVPHSSVYAHQSILRLMPLIDQYLYLSR
TOSTGCTGTOQLLPSOAGPANMSAOKNWLPGPCYRQORVSTTLSONNNFAWTGA
TKYHLNGRDSLNPVGNATHKDDERFPSPGVLMFKGQAGKNDVDYSSVMLTSEE
ELKTNVATEQYGVVADNLQNTGAPVIGTVNSQALPGMWQRNDRVYLGQPIWAKI
PHTDNFHPSYQMGFGKLPPLVKNTPVPADPTTFQAKLASITFYSTGQVS
VEIWELOKENSKNWPIQTSNTYKSTNVDPAVNTGTYSEPRPIGTYLTRLN"
560 a 667 c 568 g 422 t

BASE COUNT
ORIGIN

Alignment Scores:
Pred. No.: 2,65e-206 Length: 2217
Score: 3438.00 Matches: 624
Percent Similarity: 91.06% Conservative: 48
Best Local Similarity: 84.55% Mismatches: 64
Query Match: 86.19% Indels: 2
DB: 14 Gaps: 2

US-09-807-802A-13 (1-736) x AY243007 (1-2217)

Qy 1 MetAlaAlaAspGlyTyrLeuProAspTyrLeuGluAspAsnLeuSerGluGlyLeuArg 20
Db 1 ATGGCTGCGGATGTTATCTTCAGATTGGCTTGAGGACAACTCTCTGAGGCGCAATCGC 60
Qy 21 GluTyrTrpAspLeuLysProGlyAlaProLysProLysAlaAsnGlnLysGlnAsp 40
Db 61 GAGTGTGGGACCTGAACCTGGAGCCCGCAACCAAGCCAAACAGCAAGCAAGCAGGAC 120
Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
Db 121 GACGGCCGGGGTCTGGTGTCTCTGGCTACAGTACCTCGGACCTTCAACGGATCGAC 180
Qy 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
Db 181 AAGGGGAGCCCGTCAACCGCGGAGCGAGCGGGCCCTCGAGCAGCAAGCCCTACGAC 240
Qy 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
Db 241 CAGCAGCTCAAGCGGGTGACAATCGTACCTCGGCTAACAAGTAAACCAAGCCGCGCGAGTTT 300
Qy 101 GlnGluArgGlnGlnLysPheGlyGlyAsnLeuGlyArgAlaValPheGln 120
Db 301 CAGGAGCGGTCTCAAGAAAGATACGTTCTTGGGGGCAACCTCGGGCGAGCAGTCTTCCAG 360
Qy 121 AlaLysLysArgValLeuGluProLysGlyLeuValGluGluGlyValAlaLysThrAlaPro 140
Db 361 GCCAGAGAGGGGTTCTCGAACCTCTCGGTCTGGTTGAGGAGGCGGCTAAGACGGCTCTCT 420
Qy 141 GlyLysLysArgProValGluGlnSerProGln--GluProAspSerSerSerGlyLeu 159

Db 421 GGAAGAAGAGAGACGGGTAGAGCCATCACCTCAGCGTTCCCGGACCTCTCCACGGGCATC 480
Qy 160 GlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSer 179
Db 481 GGCAGAAGAGCCACCGAGCCCGGAGAAAGAGACTGAACATTTGGGCGAGACTGGCGACTCG 540
Qy 180 GluSerValProAspProGlnProLysGlyGluProAlaThrProAlaAlaValGly 199
Db 541 GAGTCAGTCCCGGACCTCAACCAATCGGAGAACCCACGAGCGGCCCTCTGGTCTGGGA 600
Qy 200 ProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnGluGlyAlaAsp 219
Db 601 TCTGGTACAATGGCTGAGCGCGGTGGCGCTCAATGCGACACAATAACGAAGCGCGCAC 660
Qy 220 GlyValGlyAsnAlaSerGlyAsnTyrHisCysAspSerThrTyrLeuGlyAspArgVal 239
Db 661 GAGTGGGTAGTTCTCCAGGAAATTTGGCATTTGGCATTTCCATGCTGGCGGAGCAGAGTC 720
Qy 240 IleThrThrSerThrArgThrTyrAlaLeuProThrTyrAsnAsnHisLeuTyrLysGln 259
Db 721 ATCACCAACGAGCAGCCCGAACCCTGGGCCCTGCCACCTACAACAACCATCTCTACAAGCAA 780
Qy 260 IleSerSerAlaSerThrGlyAlaSer---AsnAspAsnHisTyrPheGlyTyrSerThr 278
Db 781 ATCTCCAACGGGACATCGGAGGAGGAGCACTAAACGACAAACCTACTTTGGCTACAGCACC 840
Qy 279 ProTyrGlyTyrPheAspPheAsnArgPheHisCysHisAspSerProArgAspTyrGln 298
Db 841 CCCTGGGGGTATTTTGATTTCAACAGATTCACCTGCGACCTTCTCACCACGTGACTGGCAG 900
Qy 299 ArgLeuIleAsnAsnAsnTyrGlyPheArgProLysArgLeuAsnPheLeuLeuPheAsn 318
Db 901 CGACTCATCAACATAACTGGGGATTCCGGCCCAAGAGAGCTCAACTTCAAGCTCTTCAAC 960
Qy 319 IleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThr 338
Db 961 ATCCAGGTCAAGGAGGTCAAGCAGATGAAGGACCAAGGACCATCGGCAATAACTTACC 1020
Qy 339 SerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAla 358
Db 1021 AGCAGGATTCAGGTGTTTACGGACTCGGAATACAGCTCCGCTACGTCCTCCGCTCTGCG 1080
Qy 359 HisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyr 378
Db 1081 CACCAGGGCTGCTCCCTCCGTTCCCGCGGAGCGTCTTCATGATTTCTCAGTACGGGTAT 1140
Qy 379 LeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyr 398
Db 1141 CTGACCTTAAACAATGGCAGTCAAGGCTGTGGGCGGTTCCTCTCTTCTACTGCTGGAATAC 1200
Qy 399 PheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGlu 418
Db 1201 TTCCCTCTCAATGCTGAGGACGGGCAACAACTTTGAAATTCAGCTACACTTTCGAGGAC 1260
Qy 419 ValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeu 438
Db 1261 GTGCTTTTCCACAGCAGCTACGCGCACAGCAGGCTGGAGCGCTGAGTGAACCTCTCTC 1320
Qy 439 IleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsn 458
Db 1321 ATCAGCAGTACCTGTATTTACTTATCCAGAACTCAGTCCACGAGGAGAACTCAAGGTACT 1380
Qy 459 LysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTyr 478
Db 1381 CAGCAATTTGTTATTTTCTCAAGCGGGCCCGCAACATGTCTGGCTCAGCGCCAGACTGG 1440
Qy 479 LeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsn 498
Db 1441 CTACTGAGCGGTGTACCGTCAAGCAAGAGTCTCCAGCACACTGTCTCGAAAAACAACAC 1500
Qy 499 SerAsnPheThrThrThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIle 518

```
Db 1501 AGCAATTTTGTGGACCGGTCGCCACCAAGTATCACCTGAATGCGACGAGACTCCCTCGTT 1560
Qy 519 AenProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSer 538
Db 1561 ANTCGGCGTTCATGCTACCCACAGACGACGAGGAGCGCTCTTCCCTCAAGC 1620
Qy 539 GlyValMetIlePheGlyLysSerAlaGlyAlaSerAenThrAlaLeuAspAsnVal 558
Db 1621 GGAGTTCTAATGTTGGACGAGCGGGCTGGAAGAGCAATGTGGACTACAGCAGCGTG 1680
Qy 559 MetIleThrAspGluGluIleLysAlaThrAenProValAlaThrGluAraPheGly 578
Db 1681 ATGCTCACCGAGAGAAAGAAATAAATACTACTAACCCAGTGGCTACAGACGATGTGT 1740
Qy 579 ThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAla 598
Db 1741 GTGGTGGGACAGACACCTGACAGACACCAACCGAGCTCCCATTTGGGAACGTGTCAACAGC 1800
Qy 599 MetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIle 618
Db 1801 CAGGGGCGCTTACCTGTGTGTCGTCGCAAAACCGGACGCTGTACTCGAGGGGCCCATC 1860
Qy 619 TrpAlaValIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPhe 638
Db 1861 TGGGCCAAATTCCTCACAGGACCGCACTTTCATCTTCGCCCTGATGGGAGGCTTT 1920
Qy 639 GlyLeuLysAenProProGlnIleLeuIleLysAenThrProValProAlaAsnPro 658
Db 1921 GGACTGAAACACCGGCTCTCCAGATCTGGTGAAGAAACACTCTCTGTTCTCGGATCCT 1980
Qy 659 ProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGln 678
Db 1981 CCGACCACTTCAGCGAGCGCCCAAGCTGCTCTTTTATCAGCAGTACAGACCGGACAG 2040
Qy 679 ValSerValGluIleGluTrpGluLeuGlnLysGluAenSerLysAenTrpAsnProGlu 698
Db 2041 GTCAGGTGGGAATCGAATGGAGCTGCAGAAAGAAACAGCAGCGCTGGAACCCAGG 2100
Qy 699 ValGlnTyrThrSerAsnTyrAlaLysSerAlaAenValAspPheThrValAspAsnAen 718
Db 2101 ATTCAGTATACCTTCAACTACTACAAATCTACAAATGTGGACTTGTGCTCAATCTAG 2160
Qy 719 GlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 2161 GGTACTTATTCAGAGCTCGCCCATTTGGCACTCTGTATCTCACCGGTAATCTG 2214

RESULT 10
LOCUS AY243006 2187 bp DNA linear VRL 14-MAY-2003
DEFINITION Non-human primate Adeno-associated virus isolate AAVrh.22 capsid
protein (VP1) gene, complete cds.
ACCESSION AY243006
VERSION AY243006.1 GI:29650507
KEYWORDS
SOURCE Non-human primate Adeno-associated virus
ORGANISM Non-human primate Adeno-associated virus
VIRUSES; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
REFERENCE 1 (bases 1 to 2187)
AUTHORS Gao, G., Alvira, M.R., Somanathan, S., Lu, Y., Vandenberghe, L.H.,
Rux, J.J., Calcedo, R., Sammiquel, J., Abbas, Z. and Wilson, J.M.
TITLE Adeno-associated viruses undergo substantial evolution in primates
during natural infections
Proc. Natl. Acad. Sci. U.S.A. 100 (10), 6081-6086 (2003)
JOURNAL 12716974
PUBMED 2 (bases 1 to 2187)
AUTHORS Gao, G., Alvira, M.R., Somanathan, S., Lu, Y., Vandenberghe, L.H.,
Rux, J.J., Calcedo, R., Sammiquel, J., Abbas, Z. and Wilson, J.M.
DIRECT SUBMISSION
TITLE Submitted (23-FEB-2003) Department of Medicine, University of
Pennsylvania School of Medicine and Wistar Institute, 204 Wistar
Institute, 3601 Spruce Street, Philadelphia, PA 19104, USA
JOURNAL
FEATURES
Location/Qualifiers
source 1..2187
```

```
/organism="Non-human primate Adeno-associated virus"
/proviral
/mol_type="genomic DNA"
/isoates="AAVrh.22"
/db_xref="taxon:226582"
1..2187
/genes="VP1"
/genes="VP1"
/codon_start=1
/product="capsid protein"
/protein_id="AA088192.1"
/db_xref="GI:29650508"
/translation="MAADGYLPDMLNLSSEGREWDLKPGAPKPKANQKQDDGRG
LVLPYKYLDPFNGLDGFEVNAADAALAEHDKAYDQOLKAGDNPYLYRYHADABFQ
RLQEDTSFGGNLGRAVFOAKRVLGELVEGAKTAPGKRPDIESPDSTGKGGQ
OPAKKLNFGQGTDSSESDPDPQPIGEPGAPSGSGTMAAGGAMADNNEGADGVG
NAGSNWCHDSWLDGAVITSTRTMALPTNNHLYKQISSQSGATNDNHPGVSTPMG
YDFNDFCHFSRDRWRLNNWGRPRKLFKNIQVKEVTINDGVTTIANLTS
TIQVDSSEYQLPYVLGSAHQCLPPFPADVPMIPQGYLTNNNGSOSVGRSSFCLE
YFESQMLRTGNNEFEFTFEVPHFSYAHQSQSLDRMLNPLIDQYLYLTARTOSTGS
TRQLQHQAGNTMAEQSKNWLPGPCVRQRLSKDIDSNNSNFMFTATKYHLNGRN
SLTNPGVAMATNKDDQPPINGLVFGKTAANKTTLNVLMTSEELKTTNPVAT
EYGVVSSNLQSTAGTQGTQVNSQCALPGMWQNRDVLQGTWAKIPTHDTGNFHS
PLMGGLKHPPEQLILIKNTPPANPEVTPAKFASFIQYSTGQVSVIEWELQKE
NSKRWPEIQYTSNYSKNNVFAVNNVEGVTTEPRPIGTRILTRNL"
BASE COUNT 556 a 667 c 560 g 404 t
ORIGIN
Alignment Scores:
Pred. No.: 4.64e-206 Length: 2187
Score: 3434.00 Matches: 628
Percent Similarity: 91.85% Conservative: 48
Best Local Similarity: 85.33% Mismatches: 52
Query Match: 86.09% Indels: 8
DB: 14 Gaps: 4
US-09-807-802A-13 (1-736) x AY243006 (1-2187)
Qy 1 MetAlaAlaAspGlyTyrLeuProAspTrpLeuGluAspAenLeuSerGluGlyIleArg 20
Db 1 ATGGCTCCGATGTTATCTCCAGATTGGCTCGAGGACCAACTCTCTGAGGGCATTCGC 60
Qy 21 GluTrpTrpAspLeuLysProGlyAlaProLysProLysAlaAenGlnGlnLysGlnAsp 40
Db 61 GAGTGGTGGGACTTGAACCTGGAGCCCGCAACCCAAAGCCCAACGACGAGGAC 120
Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
Db 121 GACGCGCGGGTCTGGTGTCTTCTCGGTCTACAGTACTCTCGGACCTTCAACGGACTCGAC 180
Qy 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHiAspLysAlaTyrAsp 80
Db 181 AAGGAGAGCGCGTCAACGCGCGGACGAGCGCCCTCGAGCACGACGACGACGACGAC 240
Qy 81 GlnGlnLysAlaGlyAspAenProTyrLeuArgTyrAenHisAlaAspAlaGluPhe 100
Db 241 CAGCAGCTCAAGCGGGTGCAATCCGTACTCTGGGTATACACCGCGCGGAGGAGTTT 300
Qy 101 GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAenLeuGlyArgAlaValPheGln 120
Db 301 CAGGAGCGTCTTCAAGAGATACGTCCTTTTGGGGGCAACCTCGGCGGAGCAGTCTTCAG 360
Qy 121 AlaLysLysArgValLeuGluProLeuGluValGluGluGlyAlaLysAlaLysAlaPro 140
Db 361 GCCAAGAAGCGGGTCTCGAACCTCTCGTCTGGTGGAGGAGCGCTTAAGCGGCTCCT 420
Qy 141 GlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSerGlyIleGly 160
Db 421 GGAAGAAGAGAGACCATAGAA-----TCCCGGACTCTTCACCGGACATCGGC 468
Qy 161 LysThrGlyGlnGlnProAlaLysArgLeuAenPheGlnThrGlyAspSerGlu 180
```

469 AAGAAAGCCAGCAGCCCGCTAAAAAGAGCTCAACTTTGGCGACACTGGCGACTCAGAG 528
Db
181 SerValProAspProGlnProLeuGlyGluProProAlaIleThrProAlaAlaValGlyPro 200
Qy
529 TCAGTGGCCGAGCTCAACCAATCGAGAACCCCGCGAGCGCCCTCTGGTCTGGGATCT 588
Db
201 ThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGly 220
Qy
589 GGTACATAGCTGAGCGGCTGGCGCTCCATGGCAGACATTAACAGAGGCGCGCAGGA 648
Db
221 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIle 240
Qy
649 GTGGGTAAATGCTCCGGAATTTGGCAATTCGCAATTCACATGCTGGCGCAGAGTCAATC 708
Db
241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrrAsnAsnHisLeuTyrylGlnIle 260
Qy
709 ACCACAGCAGCCCGCAGCTGGGCCCTTGGCCACCTTACCAACCACTCTTACAGCAGATA 768
Db
261 SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrrPheGlyTyrrSerThrProTrp 280
Qy
769 TCAGTTCAGAGC--GGGGCTACCAACGACCAACCACTTCTTCGGCTACAGCACCCCTGG 825
Db
281 GlyTyrrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu 300
Qy
826 GGTATTTTGAATTCACAGATTCACATGCTGCTTCTCACCAGCTGAGCTGGCAGGACTC 885
Db
301 IleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLeuPheAsnIleGln 320
Qy
886 ATCAACAACACTGGGGATTCGGGCCCGAGAAAGCTCGCGTTCAANGTTGTTCACATCCAG 945
Db
321 ValLysGluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeuThrSerThr 340
Qy
946 GTCAGGAGGTACAGCAGCAACGCGCGTTACGACCATCGCTAATACTTACAGCAGC 1005
Db
341 ValGlnValPheSerAspSerGluTyrrGlnLeuProTyrrValLeuGlySerAlaHisGln 360
Qy
1006 ATTCAGGTCTTCGAGCTCGAGTACCAACTACGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1065
Db
361 GlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrrGlyTyrrLeuThr 380
Qy
1066 GGTGCTCTCCCTCGGTCTCGGAGCTGTTCAATGTTCTCAGTACGAGTATCTGACT 1125
Db
381 LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrrCysLeuGluTyrrPhePro 400
Qy
1126 CTAACAACGCGCAGTCAGTCTGTGGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1185
Db
401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrrThrPheGluGluValPro 420
Qy
1186 TCTCAGATCTGAGAACGGGCAATTAATTTGAATTCAGCTACACCTTTGAGGAGTGGCT 1245
Db
421 PheHisSerTyrrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
Qy
1246 TTCCACAGAGCTATCGGCAGCAGCCGAGCCCTGGACCGCTGATGATCCCTCTATCGAC 1305
Db
441 GlnTyrrLeuTyrrTyrrLeuAsnArgThrGlnAsnGlnSerClySerAlaGlnAsnLysAsp 460
Qy
1306 CAGTACTCTGACTACTCTGCGCGGACCCAGAGCACTACGCGGTCC-----ACAAGGGAG 1359
Db
461 LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuPro 480
Qy
1360 CTGCACTTCATCAGCTGGGCCCAACACATGCGCGAGCAATCAAGAACTGGTGGCC 1419
Db
481 GlyProCysTyrrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsn 500
Qy
1420 GGNCCCTGTATCGCGCGGAGAGCTGTCTAAGAGACATAGACAGCAACCAACACAGTAAAC 1479
Db
501 PheThrTrpThrGlyAlaSerLysTyrrAsnLeuAsnGlyArgGlnSerIleLeuAsnPro 520
Qy
1480 TTTCCTGGACCGGGGCCAATAATACCATCTGATGTTAGAAATTCATTAACCAACCGG 1539
Db
521 GlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyVal 540
Qy
1540 GGCCTAGCCATGGCCCAACAGGAGCAGGAGGACCACTGTTCTTTCCCATCAACCGAGTG 1599
Db

541 MetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIle 560
Qy
1600 CTGTTTTTTTGGCAAAACCGGGCT----GCCAACAGAACACGCTGGAAACAGCTGCTAATG 1656
Db
561 ThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrVal 580
Qy
1657 ACCAGCGAGGAGGAGATCAAAACCCACCAATCCCGTGGCTACAGAAAGATACGGTGTGGTC 1716
Db
581 AlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGly 600
Qy
1717 TCCAGCAACCTGCAATCGTCTACGGCCGAGCCCCAGACACAGACTGTCAACAGCCAGGGG 1776
Db
601 AlaLeuProGlyMetValTrpGlnAspArgAspValTyrrLeuGlnGlyProIleTrpAla 620
Qy
1777 GCTCTGCCCGGCACTGCTTGGCAGAACCGGAGCGTGTACCTGAGGGTCCCATCTGGGCC 1836
Db
621 LysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeu 640
Qy
1837 AAAATTCCTCACCGGACGCACTTTTCCCGCTCTCCCTGATGGCGGATTTGGACTC 1896
Db
641 LysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAla 660
Qy
1897 AAACACCGGCTCTCTCAATTTCTCATCAAAACACCCCGGTACCTGTCTAATCTCCAGAG 1956
Db
661 GluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrrSerThrGlyGlnValSer 680
Qy
1957 GTGTTTACTCTCGCAAGTTTGCCTCATTTATCAGCAGTACAGCACCGGCGCAGGTCCAGC 2016
Db
681 ValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGln 700
Qy
2017 GTGGAGATCGAGTGGGAATTCAGAAAGAGAACAGCAACGCTGGAATCCAGAGATTCAG 2076
Db
701 TyrrThrSerAsnTyrrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeu 720
Qy
2077 TACACCTCAATATTATGCAAGTCTAATAATGTGGAATTTGCTGTCAACAAGAGGGGTT 2136
Db
721 TyrrThrGluProArgProIleGlyThrArgTyrrLeuThrArgProLeu 736
Qy
2137 TATACTGAGCTCGCCCATTTGGCACCGCTTACCTACCCGTAACCTG 2184
Db

RESULT 11

AY243009 2217 bp DNA linear VRL 14-MAY-2003
LOCUS
DEFINITION Non-human primate Adeno-associated virus isolate AAVrh.18 capsid
protein (VP1) gene, complete cds.
ACCESSION AY243009
VERSION AY243009.1 GI:29650513
KEYWORDS
SOURCE Non-human primate Adeno-associated virus
ORGANISM Non-human primate Adeno-associated virus
REFERENCE 1 (bases 1 to 2217)
AUTHORS Gao, G., Alvira, M.R., Somanathan, S., Lu, Y., Vandenbergh, L.H.,
Rux, J., Calcedo, R., Sammiguel, J., Abbas, Z., and Wilson, J.M.
TITLE Adeno-associated viruses undergo substantial evolution in primates
during natural infections
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (10), 6081-6086 (2003)
PUBMED 12716974
REFERENCE 2 (bases 1 to 2217)
AUTHORS Gao, G., Alvira, M.R., Somanathan, S., Lu, Y., Vandenbergh, L.H.,
Rux, J., Calcedo, R., Sammiguel, J., Abbas, Z., and Wilson, J.M.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-2003) Department of Medicine, University of
Pennsylvania School of Medicine and Wistar Institute, 204 Wistar
Institute, 3601 Spruce Street, Philadelphia, PA 19104, USA
FEATURES
source location/Qualifiers
1. .2217
/organism="Non-human primate Adeno-associated virus"
/proviral
/mol_type="genomic DNA"
/isolate="AAVrh.18"
/db_xref="taxon:226592"

```

gene      1. .2217
CDS       /gene="Vp1"
          1. .2217
          /gene="Vp1"
          /codon_start=1
          /product="capsid protein"
          /protein_id="RAO88195.1"
          /db_xref="GI:29650514"
          /translation="MAADGYLPDWLEEDNLSEGI REWMDLKPQPKPKANQOQDDRG
          LVLPGYKYLGFNGLDGEPVNEADAALBHDKAYDKOLEQDNPYIKYNHADAFOE
          KLOEDTSGFGLNGLRAVFAKQKRVLEPLGLVEEGATAPGKRPVPEPSPPSDSTGI
          RGTQOPAKKRLNFQCTGDSVDPDPQPIGRPPAGPSGLSGSTWAAAGSGAPWADNNEG
          ADGVSSSGNHCDSLWLDGVITISTRTWALPTYNHLYKQISNGTSGSTNDNTYF
          IANNLTSTIQTFTSEYQLPYVLGSAHGCCLPPFPADVFMIPOGYLTLLNNGSAVGR
          SSFYCLETFPQMLRTGNFVSQFEDVPFSSYAHQSQSLDRMLNPLIDLYLYLSR
          TQSTGTGATQMLLPFOAGPNNSAAQKWLPGPCYRQORVSTTLLSQNNNSFAWTGA
          TKYHLNGRDSLVNPGVAMATHKDEERFPSSGVLMFGKQAGKNDVDYSVMLTSEB
          EIKTNPVATQYGVVADNLQOQNAAPIVGAVNSQGLPFGMWQNRDYLQGPINAKI
          PHTDNFPSPMGFGUKHPFPQLILNTPVPADPPTTFSQAKLASITQISTGQVS
          VEIENLQKNSKRNPEIQYTSNYKSTNVDFAVNTEGTYSEPRPIGTRYLTRNL"
BASE COUNT 563 a 673 c 568 g 413 t
ORIGIN

Alignment Scores:
Pred. No.: 6,3e-206 Length: 2217
Score: 3432.00 Matches: 623
Percent Similarity: 90.92% Conservative: 48
Best Local Similarity: 84.42% Mismatches: 65
Query Match: 86.04% Indels: 2
DB: 14 Gaps: 2

US-09-807-802A-13 (1-736) x AY243009 (1-2217)

Qy 1 MetAlaAlaAseGlyTyrLeuProAspTyrLeuGluAseAseLeuSerGluGlyIleArg 20
Db 1 ATGGCTCGCATGGTATCTTCAGATTGGCTCGAGGACACCTCTCTGAGGGCATTCGC 60

Qy 21 GluTyrPAspLeuLysProGlyAlaProLysProLysAlaAseGlnGlnLysGlnAse 40
Db 61 GAGTGGTGGACTTGAACCTGGAGCCCGCAAAACCCAAAGCCAAAGCAGGAC 120

Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAseGlyLeuAse 60
Db 121 GAGCGCCGGGGTCTGGTCTCTCTCGCTACAGTACCTCGACACCTTCAACGGACTCGAC 180

Qy 61 LysGlyGluProValAsnAlaAlaAseAlaAlaLeuGluHisAseLysAlaTyrAse 80
Db 181 AAGGAGAGCCGGTCAACGAGGACAGCGCGCGGCTCGAGCAGCAGAGGCTTACGAC 240

Qy 81 GlnGlnLeuLysAlaGlyAseAseProTyrLeuArgTyrAseHisAlaAseAlaGluPhe 100
Db 241 AAGCAGCTCGAGCAGGGGACAAACCGTACCTCAAGTACCTCGACACCTTCAACGGACTCGAC 300

Qy 101 GlnGluArgGlnGlnAsePheSerPheGlyGlyAseLeuGlyArgAlaValPheGln 120
Db 301 CAGGAGCGTCTTCAAGAGATACGTCTTTTGGGGGCAACCTCGGGGAGCAGTCTTCCAG 360

Qy 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaPro 140
Db 361 GCCAAGAGCGGGTCTCGAACCTCTCGGTCTGGTGTAGGAGGCGCTAAGCGGCTCTCT 420

Qy 141 GlyLysLysArgProValGluGlnSerProGln---GluProAsePseSerSerGlyIle 159
Db 421 GGAAAGAGAGAGACCGGTAGAGCATCACCCAGCGCTTCTCCAGACTCTCTACGGGCATC 480

Qy 160 GlyLysThrGlyGlnProAlaLysLysArgLeuAsePheGlyGlnThrGlyAseSer 179
Db 481 GGCAAGAGAGAGACCGGTAGAGCATCACCCAGCGCTTCTCCAGACTCTCTACGGGCATC 540

Qy 180 GluSerValProAseProGlnProLeuGlyGluProProAlaThrProAlaAlaValGly 199
Db 541 GAGTCAGTCCCGGACCTTCAACCAATCGAGAGAACCCCGCAGGCGCTCTCTGGTCTGGA 600

```



```
Qy 220 GlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgVal 239
Db 661 GGAGTGGGTATCTTCAGGAAATGGCGATTGCGATTCCACATGCTGGGGCAGAGATC 720
Qy 240 IleThrThrSerArgThrTrpAlaLeuProThrTyAsnAsnHisLeuTyLysGln 259
Db 721 ATCCACCACGACCCGAACTTGGCCCTCCACCTTACAAACACCACTCTTACAGCA 780
Qy 260 IleSerSerAlaSerThrGlyAlaSer---AsnAspAsnHisTyPheGlyTySerThr 278
Db 781 ATCTCCACGGGACTTCGGGAGGAGCACCACGACCAACACTACTTCTGGCTACGAC 840
Qy 279 ProTrpGlyTyPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGln 298
Db 841 CCCTGGGGGTATTTGACTTTAAACAGATTCCACTGCCACTTCTCACCACGTCGAC 900
Qy 299 ArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLeuPheAsn 318
Db 901 CGACTCATCAACAACTTGGGGATTCGGGCCCAAGAGACTCAACTTCAAGCTTTCA 960
Qy 319 IleGlnValLysGluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeuThr 338
Db 961 ATCCAGGTCAAGGAGTCCAGCAGAAATGAAGCCACCAAGACCATCGCCAAATA 1020
Qy 339 SerThrValGlnValPheSerAspSerGlyTyGlnLeuProTyValLeuGlySerAla 358
Db 1021 AGCAGATTCCAGGTCTTTACGGACTCGGAATACCAAGTCCCGTACGCTCCGGCT 1080
Qy 359 HisGlnGlyCysLeuProPheProAlaAspValPheMetIleProGlnTyGlyTy 378
Db 1081 CGCCAGGCTGCTTCCGTCCTCCGCGGAGCTTTCATGATTCCTCAGTACGGGTAC 1140
Qy 379 LeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyCysLeuGluTy 398
Db 1141 CTGACTCTGAACAATAGGCGAGTCAGCGCCGCGGCGCTTCTCTCTACTGCTGGAGTAC 1200
Qy 399 PheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyThrPheGluGlu 418
Db 1201 TTTCTCTCTCAAAATCTGAGAACGGGCAACAACTTTGAGTTGAGTACCACTTGA 1260
Qy 419 ValProPheHisSerSerTyAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeu 438
Db 1261 GTGCTTTTCAGCAGCTACGCGCACAGCCAAAGCTTGGACCGCTCATGAACCCCTC 1320
Qy 439 IleAspGlnTyLeuTyTyLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsn 458
Db 1321 ATCGACCAAGTACCTGTACTTACCTGTCTCGGACTCAGTCCACGGGAGGTACCG 1380
Qy 459 LysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrp 478
Db 1381 CAGCAGTTGCTATTTCTCAGCGCCGGGCTTAATAACATGTCGGCTCAGGCCCAAA 1440
Qy 479 LeuProGlyProCysTyArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 498
Db 1441 CTACCCGGGCGCTGTACCGGCAGCAACCGCTCTCCACGACACTGTCCGCAAAATA 1500
Qy 499 SerAsnPheThrTrpThrGlyAlaSerLysTyAsnLeuAsnGlyArgGluSerIle 518
Db 1501 AGCAACTTTGCTGACCGGTGCCCAAGATATCATCTGAATGGCAGAGACTCTCTG 1560
Qy 519 AsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSer 538
Db 1561 AATCCGGGTGCTGATGGCAACCCACCAAGGACGACGAGAGCGAATTTTTCCGTC 1620
Qy 539 GlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnVal 558
Db 1621 GGAGTCTTAATGTTTGGGAAACAGGAGCTGGAAGGCAACGCTGACTATAGCAGCT 1680
Qy 559 MetIleThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGly 578
Db 1681 ATGCTAACCAACGTCAGGAAGAAATATAAACCAACCCAGTGGCCACAGAACGATCG 1740
Qy 579 ThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAla 598
```

```
Db 1741 GTGTTGGCCGATACCTGCAACAGCAAAACCGCGCTCTATTATTAGGGCGTCAACAGT 1800
Qy 599 MetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyLeuGlnGlyProIle 618
Db 1801 CAAGGAGCCTTACTTGGCATGGTCTGGCAGAACCGGACGCTGTACCTGCAGGGTCTTATC 1860
Qy 619 TrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPhe 638
Db 1861 TGGGCCCAAGATTCTCACAACGACGGAACATTTTCATCCCTCGCGCTGTATGGGGGCTTT 1920
Qy 639 GlyLeuLysAsnProProGlnIleLeuLysAsnThrProValProAlaAsnPro 658
Db 1921 GGACTGAAACACCGCGCTCTCTAGATCTCTGATTAAAGAAATACACTGTTCCTCCGGATCCT 1980
Qy 659 ProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTySerThrGlyGln 678
Db 1981 CCAACTACCTTCAGTCAAGCTAAGTGGCGTCTTCATCAGCAGTACACACCGGACAG 2040
Qy 679 ValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGlu 698
Db 2041 GTCAGCGTGGAAATTGAATGGGAGCTGCAGAAAGAAACAGCAACCGCTGGAACCCAGAG 2100
Qy 699 ValGlnTyThrSerAsnTyAlaLysSerAlaAsnValAspPheThrValAspAsnAsn 718
Db 2101 ATTCAATACACTTCCAACTACTCAAACTACAAATGTGGACTTTGCTGTTAACACAGAT 2160
Qy 719 GlyLeuTyThrGluProArgProIleGlyThrArgTyLeuThrArgProLeu 736
Db 2161 GGCATTAATTCTGAGCTCGCCCATCGGACCGTACCTCACCCTGAATCTG 2214

RESULT 13
AY243020 2214 bp DNA linear VRL 14-MAY-2003
LOCUS Non-human primate Adeno-associated virus isolate AAVcy.2 capsid
DEFINITION protein (VP1) gene, complete cds.
ACCESSION AY243020
VERSION AY243020.1 GI:29650535
KEYWORDS Non-human primate Adeno-associated virus
SOURCE Non-human primate Adeno-associated virus
ORGANISM Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
REFERENCE 1 (bases 1 to 2214)
AUTHORS Gao, G., Alvira, M.R., Somanathan, S., Lu, Y., Vandenbergh, L.H.,
Rux, J.J., Calcedo, R., Sanmiguell, J., Abbas, Z. and Wilson, J.M.
TITLE Adeno-associated viruses undergo substantial evolution in primates
during natural infections
Proc. Natl. Acad. Sci. U.S.A. 100 (10), 6081-6086 (2003)
JOURNAL 12716974
PUBMED 2 (bases 1 to 2214)
REFERENCE Gao, G., Alvira, M.R., Somanathan, S., Lu, Y., Vandenbergh, L.H.,
Rux, J.J., Calcedo, R., Sanmiguell, J., Abbas, Z. and Wilson, J.M.
AUTHORS Direct Submission
TITLE Submitted (23-FEB-2003) Department of Medicine, University of
Pennsylvania School of Medicine and Wistar Institute, 204 Wistar
Institute, 3601 Spruce Street, Philadelphia, PA 19104, USA
JOURNAL
FEATURES
source
1. .2214
/organism="Non-human primate Adeno-associated virus"
/protein="capsid protein"
/mol_type="genomic DNA"
/isolate="AAVcy.2"
/db_xref="taxon:226582"
1. .2214
/gene="VP1"
1. .2214
/gene="VP1"
/codon_start=1
/product="capsid protein"
/protein_id="AA088206.1"
/db_xref="GI:29650536"
/translat="MAADGYLPDWLENDLSEGRWMDLKPFGPKXANQKQDNGRG
LVLPGYKYLPGFPNGLDKGFVNDAADAAALSHDKAYDQDLNAGDNPYLRYNHADAEPFCE
```

RLOEDTSFGNIGRAVFOAKKRVLRPLGLVBEAGAKTAPAKKRPVBPSPQSPDSSTGI
 CKKQOPARKLNFGQTDSESPDPQLGEPAPSPSSVSGSTVAAGGAPKADNNEG
 ADVGAGNCHDSITWLDGRVITSTRWALPFTNNHLYFOISSETAGSTNDNTYFG
 YSTPWGDFNRFHCHFSRDWRLLNNWPRPKKLRPKLFNIQKVEVTTNDGVTTI
 ANNLTSTIQVSDSYQLPYLGSAAHQCLPPFADVPFMIPOYGLTLNNGSQSVGRS
 SPYCLEYFPQMLRTGNPFSEFSDYVPHSSYAHQSOLRLANPLIDQYLYLART
 QSDPGTAGNRLQFQGGPSTMAOAKNMLPQPCFOORVSKTLDONNNGNFAWTGA
 TKYHLNGRSLNVPGVAMATHKDDERFPSSSVLI FGKTGATNKTTLLENVMTNNEE
 IRPTTPVATEEYGVIVSSNLOAAATAAQTOVNNQALPGMWQRDQVYLQSPIWAKIP
 HTDGNFSPKMGGLGHKHPQIILIKNTVPANPEVFTPAKFAFPIQYSTGVSV
 ETEWELGNSKRWDPETQYTSNFKQTVDPVDSQGVYSEPRPIGTRYLTRLN"

BASE COUNT 557 a 663 c 565 g 429 t

Alignment Scores:
 Pred. No.: 1,86e-205 Length: 2214
 Score: 3424.50 Matches: 626
 Percent Similarity: 91.88% Conservative: 53
 Best Local Similarity: 84.71% Mismatches: 55
 Query Match: 85.85% Indels: 5
 DB: 14 Gaps: 4

US-09-807-802A-13 (1-736) x AY243020 (1-2214)

QY 1 MetAlaAlaAspGlyTyrLeuProAspTyrLeuGluAspAsnLeuSerGluGlyIleArg 20
 DB 1 ATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGGACCACTCTCTGAGGGGCATTGCG 60
 QY 21 GluTyrTyrAspLeuLysProGlyAlaProLysProLysAlaAsnGlnGlnLysGlnAsp 40
 DB 61 GAGTGTGGGACCTGAACCTCGAGCCCGGAAACCCAAAGCCAAAGCAGGAGCAGGAC 120
 QY 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
 DB 121 AACGGCCGGGGCTGGTGCTCTCTGGCTACAAGTACTCTCGAGCCCTTCAACGGGACTCGAC 180
 QY 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHiAspLysAlaTyrAsp 80
 DB 181 AAGGGGAGCCCTCAACGGGGAGCGAGCGAGCCCTCGAGCAGCAAGCCCTACGAC 240
 QY 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHiAlaAspAlaGluPhe 100
 DB 241 CAGCAGCTCAATGCGGGTGACAAACCGGTACTCTCGGTATAACCCAGCCGACCGCGAGTTT 300
 QY 101 GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyValGluAlaValPheGln 120
 DB 301 CAGGACGCTCGCAGAGATAGCTCATTTTGGGGGCAACTCGGGGAGCAGCTCTTCCAG 360
 QY 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaPro 140
 DB 361 GCCAAGAAGCGGGTCTCGAACCTCTCGGTCTGGTTGAGGAGGCGCTAAGACGGCTCCT 420
 QY 141 GlyLysLysArgProValGluGlnSerProGln---GluProAspSerSerSerGlyIle 159
 DB 421 GCNAAAGAGAGACCGGTAGAGCGCTCACCTCAGCGTTCCCGGACTCTCCACCGGGCATC 480
 QY 160 GlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSer 179
 DB 481 GGCNAGAGGGCCAGCGCCGCGCAGAGAGACTCAATTTGGTTCAGACTCGGCNACTCA 540
 QY 180 GluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGly 199
 DB 541 GAGTCAGTCCCGACCTCAACCTCTCGGAGAACCTCCAGCAGCGCCCTCTAGTGTGGGA 600
 QY 200 ProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAsp 219
 DB 601 TCTGGTACAGTGTGCGCGCGTGGCGCACCAATGGCAGACAATAACGAAGTGGCGAC 660
 QY 220 GlyValGlyAsnAlaSerGlyAsnTyrPheHisCysAspSerThrTyrLeuGlyAspArgVal 239
 DB 661 GGAAGTGGTAAATGCCCTCAGGAATTTGGCATTTGCATTTCCATGCTGGCGCAGCAGATC 720
 QY 240 IleThrThrSerThrArgThrTyrAlaLeuProThrTyrAsnAsnHisLeuTyrLysGln 259

DB 721 ATTACCAACGAGCCCGAAGCTGGCCCTGCCACCTACCAACACCACTCTACGAGCA 780
 QY 260 IleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrPro 279
 DB 781 ATCTCCAGTGAACATGCAAGTAGTACCAACAGCAACACCTACTTTCGGGTACGACACCC 840
 QY 280 TrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTyrGlnArg 299
 DB 841 TGGGGGTATTTGACTTTTACAGATTCACTGCCCATCTCTCACCGTGTGACTGCGAGCA 900
 QY 300 LeuIleAsnAsnAsnTyrGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIle 319
 DB 901 CTCATCAACAACAACCTGGGATTCCGGCCCAAGAGCTGCGGTTCAAGCTCTTCAACATC 960
 QY 320 GlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnLeuThrSer 339
 DB 961 CAGGTCAAGGAGGTCAACGAGAAATGACGGCGTTTACGACCATCGCTTAATACTTACCA 1020
 QY 340 ThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlyLysAlaHis 359
 DB 1021 ACGATTCAAGTATTCTCGGACTCGGAATACGAGCTGCGGTCTCTCGGCTCTGGCAC 1080
 QY 360 GlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeu 379
 DB 1081 CAGGGCTGCCTGCTCCGTTCCCGCGGACGCTTCTATGATTCTCAGTACGGCTACCTG 1140
 QY 380 ThrLeuAsnAsnGlySerGlnAlaValGlyArgSerPheTyrCysLeuGluTyrPhe 399
 DB 1141 ACTCTCAACAATGCACTGCTGCTGGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1200
 QY 400 ProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluVal 419
 DB 1201 CCTCTCAGATGCTGAGAACCGGGCAACAACTTTCAGTTTTCAGTACAGCTTCGAGGAG 1260
 QY 420 ProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIle 439
 DB 1261 CCTTTCACAGCAGCTACGACACACAGCAGCAGCTGAGCCGCTGATGAATCCCTCATC 1320
 QY 440 AspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGln---SerGlySerAlaGlnAsn 458
 DB 1321 GACCACTACTTGTACTACTCTGCGCAGAACACAGAGTACCCAGGAGGACAGCTGGCAAT 1380
 QY 459 LysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTyr 478
 DB 1381 CGGGAACCTGCAGTTTACCAAGCGGGCTTCAACTATGGCCGAACCAAGCAAGATTGG 1440
 QY 479 LeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsn 498
 DB 1441 TTACCTGACCTGCTTCCGGCAACAAAGGGTCTCCAAACGCTGGATCAAAACCAACAC 1500
 QY 499 SerAsnPheThrTyrThrGlyAlaSerLysTyrAsnLeuAsnGlyValGluSerIleIle 518
 DB 1501 AGCAACTTGTGCTGAGTGTGCGCCCAACAAATATCACCTGAACGGCAGAACTCGTTGG 1560
 QY 519 AsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSer 538
 DB 1561 AATCCCGCGCTCGCCATGGCAACTCACAAAGCAGCAGGAGCAGCTTTTCCATCCAGC 1620
 QY 539 GlyValMetIlePhePheGlyLysGluSerAlaGlyAlaSerAsn---ThrAlaLeuAsp 557
 DB 1621 GGAGTCTGATTTTGGAAAA-----ACTCGAGCACTTAACAAACTTACATTGGAAAT 1674
 QY 558 ValMetIleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 577
 DB 1675 GTGTTAATGACAAATGAAGAAGAAATTCGCTCTACTAATCTCTGTAGCCACGGAAGATAC 1734
 QY 578 GlyThrValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHis 597
 DB 1735 GGGATAGTCAGCAGCAACTTACAAAGCGGCTAATACTGCGAGCCGACACACAAGTTGCAAC 1794
 QY 598 AlaMetGlyAlaLeuProGlyMetValTyrGlnAspArgAspValTyrLeuGlnGlyPro 617


```

Db 1795 AACGAGGAGCCTTACCTGGCATGCTCGGACGAACCGGAGCTGTACCTGACGGTCCC 1854
Qy 618 IletpLaLeYsIleProHiThraSpGlyHisPheHisProSerProLeuMetGlyGly 637
Db 1955 ATCTGGCCAGATTCCTCACCGGATGGCAACTTTACCCGCTCTCTTTGATGGGGCGC 1914
Qy 638 PheGlyLeuYsAenProProGlnIleLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 657
Db 1915 TTTGGACTTAAACATCCGCTCTCAGATCCTGATCAAGAACACTCCCGTTCCTCGCTAAT 1974
Qy 658 ProProAlaGluPheSerAlaThrySPheAlaSerPheAlaSerPheAlaSerThryGly 677
Db 1975 CCTCCGGAGGTGTTATCTCCGCCAAGTTTCTGCTTCATCACACATGACAGCCCGGA 2034
Qy 678 GlnValSerValGluIleGluTropGluLeuGlnYsGluAenSerLysArgTirPAsnPro 697
Db 2035 CAAGTCAGCTGGAATCAGTGGGAGCTGCAGAGGAAACAGACAGCGCTGGGACCG 2094
Qy 698 GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAenValAspPheThrValAspAen 717
Db 2095 GAGATTCACTACACCTCCAACTTTGAAAGACAGACTGGTGTGGACTTTGCCGTTGACAGC 2154
Qy 718 AsnGlyLeuYsThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 2155 CAGGTGTGTTACTCTGAGCCTCGCCCTATTGGCACTCGTTACCTCACCCGTAATCTG 2211

RESULT 14
LOCUS AV243021 2208 bp DNA linear VRL 14-MAY-2003
DEFINITION Non-human primate Adeno-associated virus isolate AAVch.5 capsid
protein (VP1) gene, complete cds.
ACCESSION AV243021
VERSION AV243021.1 GI:29650537
KEYWORDS .
SOURCE Non-human primate Adeno-associated virus
ORGANISM Non-human primate Adeno-associated virus
REFERENCE 1 (bases 1 to 2208)
AUTHORS Gao, G., Alvira, M.R., Somanathan, S., Lu, Y., Vandenbergh, L.H.,
Rux, J., Calcedo, R., Sammiguel, J., Abbas, Z., and Wilson, J.M.
TITLE Adeno-associated viruses undergo subantial evolution in primates
during natural infections
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (10), 6081-6086 (2003)
PUBMED 12716974
REFERENCE 2 (bases 1 to 2208)
AUTHORS Gao, G., Alvira, M.R., Somanathan, S., Lu, Y., Vandenbergh, L.H.,
Rux, J., Calcedo, R., Sammiguel, J., Abbas, Z., and Wilson, J.M.
DIRECT SUBMISSION
TITLE Submitted (23-FEB-2003) Department of Medicine, University of
Pennsylvania School of Medicine and Wistar Institute, 204 Wistar
Institute, 3601 Spruce Street, Philadelphia, PA 19104, USA
JOURNAL Location/Qualifiers
FEATURES
source
1..2208
/organism="Non-human primate Adeno-associated virus"
/proviral
/mol_type="genomic DNA"
/isolate="AAVch.5"
/db_xref="taxon:226582"
1..2208
/gene="VP1"
1..2208
/codon_start=1
/gene="VP1"
/product="capsid protein"
/protein_id="AAO88207.1"
/db_xref="GI:29650538"
/translation="MAADGILPDWLEDLTSRGIRQWKLKPGPPPPKPNQHRDSDRG
LVLPGLYKLPFGNLDKGEVNEADAALHDKAYDHLQKQDNPIYLYKYNHDAEFOE
RLQDTSFGNLRGAVFOAKRVLLEPLVEEATPKRPTEQSPAPDSGGIG
KSGQPAKRLNFGTGTSEVPQPIGEPPAAPSGVSNMASGCGAPMADNNEGA
DVGNGSSNWHICSDSTMGDRVITSTRWALPTYNHLYKQISSESGATNDNHVFGYS
TPWGVFDRNRHCHFSRDMQRLNNWNGFRPKLNPKLNIQVKEVTQNDGTTIANF
NLSTVQVFTDSEYQLPLVLSAHQGGCLPPFPADVFMIPQYGLTLNNGSQAVGRSSF

```

```

YCLEYPPSMLRTGNNTFTSYTFEDVPFHSSVAHSQSRLDRLMNPDLIDQVLYLSKTKQG
TSGTTCQSLQFSGQSPSSMAQKWLPGSPYRQRMGSKTANDNNSEFPAMTAATKY
YLNRSLSLVNPGPFWASHKDDSEKYPWMGNLIPGQKGTCTTNVDLSEVLITDEBBIR
TNPVATEOYGOVATNHQSONTTASVSDSGILPGMWODRDVLYQGPIMAKTPTH
DGHFPHSPMGSGFLKHPPTQILLIKTNPUPANPATFTTSGKPSASEIOTVSTGVSVEI
DEWELKENSIRWNPEIQYTSNKNKSNVVEFTVDANGVSEFPFIRIGRILYLRNL".
BASE COUNT 617 a 638 c 518 g 435 t
ORIGIN
Alignment Scores:
Pred. No.: 2,14e-205 Length: 2208
Score: 3423.50 Matches: 619
Percent Similarity: 90.22% Conservative: 45
Best Local Similarity: 84.10% Mismatches: 71
Query Match: 85.82% Indels: 1
DB: 14 Gaps: 1
US-09-807-802a-13 (1-736) x AV243021 (1-2208)
Qy 1 MetAlaAlaAspGlyTyrLeuProAspTirPLeuGluAspAenLeuSerGluGlyIleArg 20
Db 1 ATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGGACACTCTCTCTGAAGGAATCAGA 60
Qy 21 GluTirPAspLeuYsProGlyAlaProLysProLysAlaAenGlnLysGlnAsp 40
Db 61 CAGTGTGGAGGCTCAACCTGCCCCACCACCGCCGAACTTAACCAACACACCGGAC 120
Qy 41 AspGlyValGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAenGlyLeuAsp 60
Db 121 GACAGTAGGGTCTTGTCTTCTGGGTACAGTACTCTCGACCTTCAACGAGACTCGAC 180
Qy 61 LysGlyGluProValAenAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
Db 181 AAAGGAGAGCGGTCAACGAGGCGAGCGCGCGCTCGAGCACGACAAAGCTACGAC 240
Qy 81 GlnGlnLeuYsAlaGlyAspAenProTyrLeuArgTyrRanHisAlaAspAlaGluPhe 100
Db 241 CACCAGCTCAAGCAGGGGACAAACCCGTACCTCAATATACCAACACCGCGAGCGTGAATTT 300
Qy 101 GlnGluArgLeuGlnGluAspThrSerPheGlyValAenLeuGlyArgAlaValPheGln 120
Db 301 CAGGAGCGTCTTCAAGAGATACGTCTTTCGGGGGACCTCTCGGCGAGGAGCTTCCAG 360
Qy 121 AlaLysArgValLeuGluProLeuGlyLeuValGluGlyAlaLysThrAlaPro 140
Db 361 GCCAAAAGAGGCTACTCGAGCTCTTGTGTGGTTGAGGAAGCTGTTAAGACGGCTCT 420
Qy 141 GlyLysArgProValGluGlnSerProGlnGluProAspSerSerSerGlyIleGly 160
Db 421 GSAANAAGAGACTATAGAGCAGTCTCTGCGAAGACCGGACTCTTCTCTCGGCGATCGGC 480
Qy 161 LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
Db 481 AATCAGCCAGCAGCGCCGCTAGAAAAGACTCAATTTGTCAGACTGCGGACACAGAG 540
Qy 181 SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro 200
Db 541 TCAGTCCCAGACCTCAACCAATCGGAGAACCCCGGACGCCCTCTCTCTGTTGGGATCT 600
Qy 201 ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAenGluGlyAlaAspGly 220
Db 601 AATACAAATGGCTTCAGCGGTGGGGGACCAATGGCAGACAAATACCAAGCGCGGACGGA 660
Qy 221 ValGlyAsnAlaSerGlyAenTirPHisCysAspSerThrTirPLeuGlyAspArgValIle 240
Db 661 GTGGTAATTCCTCGGGAATTTGGCATTTCCAGATGGATGGGCGACAGAGTTATC 720
Qy 241 ThrThrSerThrArgThrTirPAlaLeuProThrTyrAenAenHisLeuYsGlnIle 260
Db 721 ACCACGACGACAAAGACCTGGGCGCTCCACCACCTACAAATACCTCTACAGCAATC 780
Qy 261 SerSerAlaSerThrGlyAlaSerAenAspAenHisTyrPheGlyTyrSerThrProTirP 280

```

Db 781 TCCAGCAATCG---GGAGCCACCAACGACCACTACTTCGGCTACAGCACCCCTGG 837
 Qy 281 GlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTyrGlnArgLeu 300
 Db 838 GGGTATTTTGAATTTAAACAGATTCCACTGCTCACTTCTCACCACGCTGAGCGAGGACTC 897
 Qy 301 IleAsnAsnAsnTyrGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGln 320
 Db 898 ATCAACACAACTGGGGATTTAGACCAAGAACTCAATTTCAAGCTCTTCAACATCCAA 957
 Qy 321 ValLysGluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeuThrSerThr 340
 Db 958 GTCACAGGAGTCAAGCAGAAATGATGACACAGCACCATCGCCAAATAACCTTACCAGCAG 1017
 Qy 341 ValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGln 360
 Db 1018 GTGCAGGCTTTCACAGACTCTGAGTACCAAGCTGCGCTCTCGGTTCTGGCTCACCAG 1077
 Qy 361 GlyCysLeuProPheProAlaAspValPheMetIle2ProGlnTyrGlyTyrLeuThr 380
 Db 1078 GGGTGGCTTCGGCGTTCCAGCAGACGCTTCATGATTCCTCAGTACGCGCTACTTGACT 1137
 Qy 381 LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 400
 Db 1138 CTGAACAATGGCAGCCCAAGCGGTAGGAGCTTCTTCACTCTAGTACTATTTCCTCC 1197
 Qy 401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 420
 Db 1198 TCTCAGATCTGAGGACGGGAACAACCTTCCAGCTACACACTTTTGAAGACGTGCT 1257
 Qy 421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
 Db 1258 TTCCACAGCAGCTACCGGCACAGCAGCTCGATCGCTGATGATGATCTCTCATTTGAC 1317
 Qy 441 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAsp 460
 Db 1318 CAGTACCTGTATTACCTGAGCAAACTCAGGGGTACAACTGGAACAGCAGCAATCCAGA 1377
 Qy 461 LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTyrLeuPro 480
 Db 1378 CTCAGTTACGCAAGCTGGCTGCTAGCTCATGCTCAGCAGCCCAAACTGGCTACCG 1437
 Qy 481 GlyProCysTyrArgGlnArgValSerLysThrLysThrAspAsnAsnSerAsn 500
 Db 1438 GGACCCAGCTACCGACAGCGAATGTCTAAGCGGTAAATGACACCAACACACAGTGA 1497
 Qy 501 PheThrTyrThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleLeuAsnPro 520
 Db 1498 TTTGCTTGGACTGCAGCCACCAATATTACCTGATGGAAGAAATTCCTGGTCAATCC 1557
 Qy 521 GlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyVal 540
 Db 1558 GGGCCCCAATGGCCAGTCAAGAGCAGATGAGGAAGATTTTCCCTCATGACCGGAAT 1617
 Qy 541 MetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIle 560
 Db 1618 CTCATCTTGGAAAACAGCCACAGCACTACCAATGTCGACATGGAATCAGTCTTATT 1677
 Qy 561 ThrAspGluGluLeuLysAlaThrAsnProValAlaThrGluArgPheGlyThrVal 580
 Db 1678 ACAGACGAAGAAGAAATCAGAACAACTAATCTCTGGCTACAGAACAAATACGACAGGT 1737
 Qy 581 AlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGly 600
 Db 1738 GCCACCAACCATCAGAGTCAAGACACCAACAGCTTCTATGGAAGTGGACACCGAGGA 1797
 Qy 601 AlaLeuProGlyMetValTyrGlnAspArgAspValTyrLeuGlnGlyProIleTyrPala 620
 Db 1798 ATCTTACCTGGAATGGTGGCAGGACCGGAGTGTCTATCTTCAAGTCCCATTTGGGCC 1857
 Qy 621 LysIleProHisThrAspGlyHisPheHisPheSerProLeuMetGlyGlyPheGlyLeu 640
 Db 1858 AAAAATCTCTCACAGCGGACGACATTTTCACTCTCTCCGCTCATGAGGCTTTGGACTG 1917

Qy 641 LysAsnProProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAla 660
 Db 1918 AAAACACCTCTCTCCAGATCTCTGATCAAAACACACCTGTGCCAGCAATCCCGCGACC 1977
 Qy 661 GluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSer 680
 Db 1978 ACTTTCATCTCTGGAAAGTTTGGCTTCCTTACCTACCCAGTATTCACCGGACAGGTGAGC 2037
 Qy 681 ValGluIleGluTyrGluLeuGlnLysGluAsnSerLysArgTyrAsnProGluValGln 700
 Db 2038 GTGGAATATAGAGTGGAGCTGCAGAAAGAAACAGCAACCGCTGGAACCCAGAAATTCAG 2097
 Qy 701 TyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeu 720
 Db 2098 TACACTCCAACTACACAGTCGGTGAATGTGAGTTTACCGTGGAGCGCAACCGTGT 2157
 Qy 721 TyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
 Db 2158 TATCTGAACCCCGCCCTATTGGCACTCGTTACCTTACCCGGAACCTTG 2205
 RESULT 15
 AY242998
 LOCUS
 DEFINITION
 Non-human primate Adeno-associated virus isolate AAVrh.37 capsid protein (VP1) gene, complete cds.
 ACCESSION
 AY242998
 VERSION
 AY242998.1 GI:29650491
 KEYWORDS
 SOURCE
 ORGANISM
 Non-human primate Adeno-associated virus
 Viruses: asDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
 REFERENCE
 1 (bases 1 to 2190)
 Gao, G., Alivira, M.R., Somanathan, S., Lu, Y., Vandenbergh, L.H., Rux, J.J., Calcedo, R., Sanmiguel, J., Abbas, Z. and Wilson, J.M. Adeno-associated viruses undergo substantial evolution in primates during natural infections
 TITLE
 JOURNAL
 PUBMED
 12716974
 REFERENCE
 2 (bases 1 to 2190)
 Gao, G., Alivira, M.R., Somanathan, S., Lu, Y., Vandenbergh, L.H., Rux, J.J., Calcedo, R., Sanmiguel, J., Abbas, Z. and Wilson, J.M. Direct Submision
 TITLE
 JOURNAL
 Submitted (23-FEB-2003) Department of Medicine, University of Pennsylvania School of Medicine and Wistar Institute, 204 Wistar Institute, 3601 Spruce Street, Philadelphia, PA 19104, USA
 FEATURES
 ..
 source
 location/Qualifiers
 1..2190
 /organism="Non-human primate Adeno-associated virus"
 /proviral
 /mol_type="genomic DNA"
 /isolate="AAVrh.37"
 /db_xref="taxon:226582"
 1..2190
 /gene="VP1"
 1..2190
 /codon_start=1
 /product="capsid protein"
 /protein_id="AAO88184.1"
 /db_xref="GI:29650492"
 /translation="MAADGYLPDLENDLSGIREWDLKFGAPKPKANQKQDDGRG
 LVLPGYKYLGPFGNLDKGEVPAADAAALHDKAYDQQLKAGDNPFLYRNHADAFFQE
 RIQEDTSFGNIGRAVFOAKKRVLEPLGLVEGAKTA PGKKRPI DPSSTGIGKGGQ
 QPAKKLNFPGTGSSEVPDQPLGEPAPASSVSGSGTMAAGGAPTDANNEGADGIV
 NASGNWCHDSVTWLDGRTTTRTWALPTNNHLYKISSSSGATNDNNHYGYSTPW
 GYDFNRHFCHFSPLRNNWGLRNNWGFKKLRFKNIQVKEVTNGDVTIANNT
 STVQVSDSEYQLPYLGSAGHQLPPFPADVPMTPYQYLTLLNNGSVSGSSPYCL
 STYFQSMRLTGNFNEFSYFEDVPSHSAHSQSLDLRLNPLIDQLYLYLARTQSTTG
 STRELQFHQAGNPMAEQSKWLPQCYRQORLSKVLDFNNNSFAWTAATKYHLNGR
 NSLTNPGI PMATNKDDEQFPFINGLVFGTKGANKTLENLVLTSEETKTTNPVA
 TEYGVSSNLSQSSTAGSQSTINSQGLPGMWQNRDVTYLGQPIWAKIPHTDGNFPA
 SPLMGGFGLHPHPPIQLIKNTFVPANPPEVFTFAKFSFTTQYTGQVSVELEWELQK

ENSKRWNEIQYTSNYAKSNVFAVNDGVYTEPRPIGTLYLTRLN"

| | | | | |
|---|-----------|--|-------|-------|
| BASE COUNT | 558 a | 662 c | 545 g | 425 t |
| ORIGIN | | | | |
| Alignment Scores: | | | | |
| Pred. No.: | 2,44e-205 | Length: | 2190 | |
| Score: | 3422.50 | Matches: | 624 | |
| Percent Similarity: | 91.85% | Conservative: | 52 | |
| Best Local Similarity: | 84.78% | Mismatches: | 53 | |
| Query Match: | 85.80% | Indels: | 7 | |
| DB: | 14 | Gaps: | 3 | |
| US-09-807-802A-13 (1-736) x AY242998 (1-2190) | | | | |
| QY | 1 | MetAlaAlaAspGlyTyrLeuProAspTyrLeuGluAspAsnLeuSerGluGlyIleArg | 20 | |
| DB | 1 | ATGGCTGCCGATGGTATATCTCCAGATTGGCTCGAGGACAACCTCTCTGAGGGCAATTCGC | 60 | |
| QY | 21 | GluTyrTrpAspLeuLysProGlyAlaProLysProLysAlaAsnGlnGlnLysGlnAsp | 40 | |
| DB | 61 | GAGTGGTGGGACCTGAAACCTGGAGCCCGGAAACCCAAAGCCAAACAGCAAAAGCAGGAC | 120 | |
| QY | 41 | AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp | 60 | |
| DB | 121 | GACGGCCGGGCTCTGGTCTCTCTGGCTACAAGTACCTCGGACCTTCAACGGGACTCGAC | 180 | |
| QY | 61 | LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp | 80 | |
| DB | 181 | AAGGGGAGCCGCTCAACGGCGGAGCGACGGCCCTCGAGCACGACAGGCTACGAC | 240 | |
| QY | 81 | GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe | 100 | |
| DB | 241 | CAGCAGCTCAAGCGGTGCAATCCGTACCTCGGTATTAACCAACGCGCGGCGAGTTT | 300 | |
| QY | 101 | GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln | 120 | |
| DB | 301 | CAGAGGCGCTGCAAGGAAGATACGTCAATTTGGGGGCAACCTCGGGCGAGCGTCTCCAG | 360 | |
| QY | 121 | AlaLysLysArgValLeuGluProLeuGlyLeuValGluGlyAlaLysThrAlaPro | 140 | |
| DB | 361 | GCCAAAGCGGTCTCGAACCTCTCGGTCTGGTTGAGGAGCGCTAAGCGGCTCCT | 420 | |
| QY | 141 | GlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerGlyIleGly | 160 | |
| DB | 421 | GGAAAGAGACCCATAGAC- -----TCTCCAGACTCCTCCACGGGATCGGC | 468 | |
| QY | 161 | LysThrGlyGlnGlnProAlaLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu | 180 | |
| DB | 469 | AAAAAGCCAGCAGCCCGCTAAAGAAAGCTCAATTTTGTGCTAGACTGGCGACTCAGAG | 528 | |
| QY | 181 | SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro | 200 | |
| DB | 529 | TCAGTCCCGGACCTCAACCTCTGGAGAACCTCCAGCAGCGGCTCTAGTGTGGGATCT | 588 | |
| QY | 201 | ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnGlnGlyAlaAspGly | 220 | |
| DB | 589 | GGTACAAATGGCTGACGGCGGTGGCGCCCAACCGCAGACAAATAACGAAGGTGGCGCGGA | 648 | |
| QY | 221 | ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIle | 240 | |
| DB | 649 | GTGGGTAAATGGCTCAGGAATAATGGCATTTGCCATTTCCACATGGCTGGCGCAGAGTCAATC | 708 | |
| QY | 241 | ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIle | 260 | |
| DB | 709 | ACCACGACCAAGAACTGGGCCCTCCCACTTACAAACCAACCACTCTACAAAGCAATC | 768 | |
| QY | 261 | SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrp | 280 | |
| DB | 769 | TCCAGCAGCAGCTCAGGAGCCCAATGACAAACCACTACTTTCGGCTACAGCACCCTCTGG | 828 | |
| QY | 281 | GlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu | 300 | |
| DB | 829 | GGGTATTTTGACTTTAACAGATTTCACCTGGCCACTTCTCACACGCTGACTGGCAGGACTC | 888 | |
| QY | 301 | IleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGln | 320 | |
| DB | 889 | ATCAACAACAACCTGGGATTCGGGCCCAAGAAGCTCGGTTCAGGCTCTTCAACATCCAG | 948 | |
| QY | 321 | ValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThr | 340 | |
| DB | 949 | GTCAAGGAGGTCAACAAGAAATCGCGCTCGACACCACTCGTAAATACCTTACGACACG | 1008 | |
| QY | 341 | ValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGln | 360 | |
| DB | 1009 | GTTCAGGTCTTCTCGGACTCGGAATACCAAGCTGCGGTAGCTCTCGGCTCTGGCGACCA | 1068 | |
| QY | 361 | GlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr | 380 | |
| DB | 1069 | GGCTGCTGCCCTCCGTTCCCGCGGACGCTTTCATGATTCCTCAGTACGGCTACCTGACT | 1128 | |
| QY | 381 | LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro | 400 | |
| DB | 1129 | CTGAACAACGGCAGCCCAATCGGTGGGCCCTCTCTCTACTGCTGGGAATATTTCCCC | 1188 | |
| QY | 401 | SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro | 420 | |
| DB | 1189 | TCTCAATCTCGAACCGGCAACACTTTGAGTTTCAGTTACAGCTTCGAGGACGTGCCT | 1248 | |
| QY | 421 | PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp | 440 | |
| DB | 1249 | TTCCACAGCAGCTTACCGCACAGCCAGAGCCTAGACCGGCTGATGAACCTCTCATCGAC | 1308 | |
| QY | 441 | GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAsp | 460 | |
| DB | 1309 | CAGTACTCTACTACTGCTGGCCCGGACCCAGAGCACCCAGGGTTCC-----ACCAGGAA | 1362 | |
| QY | 461 | LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuPro | 480 | |
| DB | 1363 | CTGCAATTCATCAAGCTGGGCCCAATATCTATGGCCGAGCAGCTCAAGAACTGGCTGCT | 1422 | |
| QY | 481 | GlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsn | 500 | |
| DB | 1423 | GGACCTCTGTATAGGCAACAGAGACTGTCAAAGAATCTTGAGACTTTTAAACAACAAGCAAT | 1482 | |
| QY | 501 | PheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleLeuAsnPro | 520 | |
| DB | 1483 | TTTGGCTGAGCTGCTGCCACTAAATATCTGAATGGCAGAACTCTTTTGACCAATCTCT | 1542 | |
| QY | 521 | GlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyVal | 540 | |
| DB | 1543 | GGCATTTCCCATGGCAACCAACAAGGATGATGAGGACCACTTCTTTCCCATCAACGGGTA | 1602 | |
| QY | 541 | MetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIle | 560 | |
| DB | 1603 | CTGGTTTTTGGCAAGACGGGAGCT---GCCAACAACCACTACGCTGGAAAAACGTTCTGATG | 1659 | |
| QY | 561 | ThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrVal | 580 | |
| DB | 1660 | ACGAGGAGGAGAGATCAAGACCACTAACCTCTGGCTACAGAAGATACGGTGTGGTC | 1719 | |
| QY | 581 | AlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGly | 600 | |
| DB | 1720 | TCCAGCACTCGAGTCGTCTACAGCGGGCCCTCAATCAACAGACTATCAACAGCAGGGA | 1779 | |
| QY | 601 | AlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAla | 620 | |
| DB | 1780 | GCAGTCTGCTGGCATGGTCTGGCAGAACCGGAGCGTGTATCTGAGGGGTCCCATCTGGGCC | 1839 | |
| QY | 621 | LysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeu | 640 | |
| DB | 1840 | AAATTCCTCACAGGATGGCAACTTTCACCGCTCTCTCTGATGGCGGTTTTTGGACTC | 1899 | |
| QY | 641 | LysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAla | 660 | |
| DB | 1900 | GAACACCGGCTCCACAGATCTCTGATCAAAAAACACACCTGTACCTGTCTTAATCTCTCCGGAG | 1959 | |


```

FT CDS 1..2211
FT /*tag= a
XX /product= "VP1 protein"
XX
XX WO200028061-A2.
XX
XX 18-MAY-2000.
XX
XX 02-NOV-1999; 99WO-US25694.
XX
XX 05-NOV-1998; 98US-0107114.
XX (UYPE-) UNIV PENNSYLVANIA.
XX
XX Wilson JM, Xiao W;
XX WPI; 2000-376571/32.
XX P-PSDB; AAY71167.
XX
XX Novel adeno-associated virus serotype 1 polynucleotide useful for
XX PT preparation of medicament for delivery of a transgene to a host -
XX
XX Claim 10; Page 83-87; 108pp; English.
XX
XX The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA
XX which is characterised by two inverted terminal repeats (ITR) and open
XX reading frames for rep and capsid (cap) proteins. The rep reading frame
XX encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
XX reading frame encodes three structural proteins, VP1, VP2 and VP3.
XX The AAV-1 sequence or its fragments particularly ITRs, rep and cap
XX coding regions, are useful in production of recombinant viral vectors
XX for gene delivery. These vectors can be used as gene therapy
XX vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
XX not induce the formation of neutralising antibodies specific to any
XX serotype of AAV hence is useful for transforming host cells, and in the
XX preparation of a medicament for the delivery of transgene to a host.
XX The present sequence is an AAV-1 DNA encoding a cap protein VP1 which is
XX useful in the production of recombinant viral vector for gene delivery.
XX
XX Sequence 2211 BP; 577 A; 652 C; 538 G; 444 T; 0 other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 2,376-283 Length: 2211
XX Score: 3989.00 Matches: 736
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 21 Gaps: 0
XX
XX US-09-807-802A-13 (1-736) x AAD00777 (1-2211)
XX
XX Qy 1 MetAlaAlaaspGlyTyrLeuProAaspTrrPLeuAlaaspAenLeuSerGluGlyIleArg 20
XX Db 1 ATGGCTGCCGATGGTATCTTCAGATTGGCTCGAGGACAACTCTCTGAGGCAATTCG 60
XX
XX Qy 21 GluTrrPLeuLeuysProGlyAlaProLyseProlyseAlaAenGlnLysGlnAap 40
XX Db 61 GAGTGGTGGGACTTGAACCTTGAGGCCCCAGGCCCAAGCCCAACGACGAAAGCAGGAC 120
XX
XX Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAenGlyLeuAap 60
XX Db 121 GACGGCCGGGGTCTGGTGTCTCTGGCTACAAGTACCTCGACCCCTTCAACGAGCTCGAC 180
XX
XX Qy 61 LysGlyGluProValAenAlaAlaAapAlaAlaLeuGluHisAplysAlaTyrAap 80
XX Db 181 AAGGGGAGCCCGTCAACGGGGCGGACGAGCGGCCCTCGAGCAGCAGCAAGGCTACGAC 240
XX
XX Qy 81 GlnGlnLeuLysAlaGlyAapAenProTrrPLeuArgTrrAenHisAlaAapAlaGluPhe 100
XX Db 241 CAGCAGCTCAAGCGGGTGAAATCCGTACTTGGGGTATTAACCCAGCCGACGCCGAGTTT 300
XX
XX Qy 101 GlnGluArgLeuGlnGluAapThrSerPheGlyAenLeuGlyArgAlaValPheGln 120
XX Db

```

```

Db 301 CAGGAGCGTCTGCAAGAGATACGCTCTTTGGGGGCAACCTCGGGGAGCAGCTCTTCAG 360
Qy 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGlyAlaLysThrAlaPro 140
Db 361 GCCAAGAGCGGGTCTCGAACCTCTCGGTCTGGTTGAGGAAGCGCTTAAGACGCGCTCCT 420
Qy 141 GlyLysLysArgProValGluGlnSerProGlnGluProAaspSerSerSerGlyLysGly 160
Db 421 GGAAGAAAGAACTCCGGTAGAGCAGTCGCCCAAGAGCCAGACTCTCTCTCGGGCATCGC 480
Qy 161 LysThrGlyGlnGlnProAlaLysLysArgLeuAenPheGlyGlnThrGlyAaspSerGlu 180
Db 481 AAGCAGGCCAGCAGCCCGCTTAAAGAGAGACTCAATTTTGTTCAGACTGGCGACTCAGAG 540
Qy 181 SerValProAaspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro 200
Db 541 TCAGTCCCGCATCCCAACCTCTCGGAGAACCTCCAGCAACCCCGCTGCTGTGGGAGCT 600
Qy 201 ThrThrMetAlaSerGlyGlyValAlaProMetAlaAapAenAenGluGlyAlaAapGly 220
Db 601 ACTCATATGGCTTCAGGCGGTGGCGCACCAATGGCAGACATTAACGAAGGCGCCGCGGA 660
Qy 221 ValGlyAenAlaSerGlyAenTrpHisCysAaspSerThrTrpLeuGlyAaspArgValIle 240
Db 661 GTGGGTATGCTTCAGGAAATTTGGCATTTGCATTCACATGCTGGGCGCAGAGTCAATC 720
Qy 241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAenAenHisLeuTrrLysGlnIle 260
Db 721 ACCACAGCAGCCCGCACCTGGGCTTGGCCACCTTCAATTAACCACTCTTACGAAGCAATC 780
Qy 261 SerSerAlaSerThrGlyAlaSerAenAaspAenHisTyrPheGlyTyrSerThrProTrp 280
Db 781 TCAGTGTCTTCAACGGGGCGCCAGCAACCACTACTTTCGGTTCAGACACCCCTCTGG 840
Qy 281 GlyTyrPheAaspPheAenArgPheHisCysHisPheSerProArgAapTrpGlnArgLeu 300
Db 841 GGGTATTTTGTATTTCAACAGATTTCCACTGCCACTTTTACCACGTGACTGGCAGGACTC 900
Qy 301 IleAenAenAenTrpGlyPheArgProLysArgLeuAenPheLysLeuPheAenIleGln 320
Db 901 ATCAACAACAATTTGGGATTCGGGCCCAAGAGACTCAACTTCAAACTCTTCAACATCCA 960
Qy 321 ValLysGluValThrThrAenAaspGlyValThrIleAlaAenAenLeuThrSerThr 340
Db 961 GTCAAGGAGGTCAACGACGAATGATGGCGTCACACCACTCGCTTAATACCTTACCAGCA 1020
Qy 341 ValGlnValPheSerAaspSerGluTyrGlnLeuProTrrPValLeuGlySerAlaHisGln 360
Db 1021 GTTCAAGTCTTCTCGGACTCGGAGTACCAGCTTCCGTACGCTCTCGGCTCTCGCAGC 1080
Qy 361 GlyCysLeuProProPheProAlaAapValPheMetIleProGlnTyrGlyTyrLeuThr 380
Db 1081 GGCTGCCCTCCCTCCGTTCGGGCGGAGCGTTCATGATTTCCGCAATACGGCTACCTGAGC 1140
Qy 381 LeuAenAenGlySerGlnAlaValGlyArgSerSerPheTrrCysLeuGluTyrPhePro 400
Db 1141 CTCACAAATGGCAGCAACCGCTGGGCGTTCATCTCTTTACTGCTGGAAATATTTCCTCT 1200
Qy 401 SerGlnMetLeuArgThrGlyAenAenPheThrPheSerTrrTrrPheGluGluValPro 420
Db 1201 TCTCAGATGCTGAGAACCGGCAACAACTTTTACCTTCAGCTACACCTTTGAGGAAGTGC 1260
Qy 421 PheHisSerSerTrrAlaHisSerGlnSerLeuAapArgLeuMetAenProLeuIleAap 440
Db 1261 TTCCACAGCAGCTACGCGCAGCAGCAGCGCTGGACCGGCTGATGATCTCTCTCATCGAC 1320
Qy 441 GlnTyrLeuTrrTrrLeuAenArgThrGlnAenGlnSerGlySerAlaGlnAenLysAap 460
Db 1321 CAATACCTGTTTACCTGAACAGAACTCAAAATCAGTCGGGAAGTGCCTCCCAACAGGAC 1380
Qy 461 LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAenTrpLeuPro 480
Db 1381 TTGCTGTTTAGCGGTGGGTCTCCAGCTGGCATGTCTGTTTCCAGCCCAAAACTGGGTACCT 1440

```

```

Oy 481 GlyProCysTrpArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsn 500
Db 1441 GGACCCGTGTTATCGCAGCAGCGGCTTCTAAACAAAAACAGACAAACACAGCAAT 1500
Oy 501 PheThrTrpThrGlyAlaSerLysTrpAsnLeuAsnGlyArgGluSerIleIleAsnPro 520
Db 1501 TTTACCTGGACTGGTGTCTCAAAATATAACCTCAATGGCGGTGAATCATCATCAACCCCT 1560
Oy 521 GlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyVal 540
Db 1561 GGCACGTGCTATGGCTCTACAAAGACGACGACGACGACGACGACGACGACGACGACGACG 1620
Oy 541 MetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIle 560
Db 1621 ATGATTTTGGAAAGAGAGCGCGGAGCTTCAACACACTGCTATGGCAATGTCTCATGATT 1680
Oy 561 ThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrVal 580
Db 1681 ACAGACGAAGAGGAATTTAAAGCCCACTAACCTGTGGCCACCGAAGATTGGGACCGTG 1740
Oy 581 AlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGly 600
Db 1741 GCAGTCAATTTCCAGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1800
Oy 601 AlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAla 620
Db 1801 GCATTACCTGGCATGGTGTGGCAAGATAGACAGCTGTACCTGCAGGGTCCCATTTGGGCC 1860
Oy 621 LysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeu 640
Db 1861 AAAATTCCTCACAGATGGACACTTTTCACCCGCTCTCTCTATGGCGGGTTTGGACTC 1920
Oy 641 LysAsnProProGlnIleLeuLysAsnThrProValProAlaAsnProProAla 660
Db 1921 AAGAACCGCCTCTCAGATCTCTATCAAAACACGCTGTCTCTCGAATCTCTCCGGCG 1980
Oy 661 GluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSer 680
Db 1981 GAGTTTTCAGCTACAAAGTTTGTCTTCTATCATCCCAATACTCCACAGGACAAGTGAGT 2040
Oy 681 ValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGln 700
Db 2041 GTGGAAATTTGAATGGAGCTCGAAGAAAGAAACAGCAGCGCTGGAATCCCGAAGTGCG 2100
Oy 701 TyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeu 720
Db 2101 TACATCCATTTATGCAAAATCTGCCAACGTTGATTTTACTGTGGACAAATGGACTT 2160
Oy 721 TyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 2161 TATACTGAGCCTCGCCCATTTGGCACCCTTACCTTACCCCTGCCCTG 2208

RESULT 2
AAD00772
ID AAD00772 standard; DNA; 4718 BP.
XX
XX AAD00772;
XX
XX 08-SEP-2000 (first entry)
XX
DE Adeno-associated virus serotype 1 DNA.
XX
KW Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
KW cap protein; recombinant viral vector; gene delivery; gene therapy;
KW vaccine; transgene; ss.
XX
OS Adeno associated virus serotype 1.
XX
FH Key Location/Qualifiers
FH repeat_unit 1..143
FT /*tag= a
FT /label= 5'_ITR
FT

```

/note= "Inverted terminal repeat which is capable of forming T-shaped hairpin structure"
 89..110
 /*tag= b
 /bound moiety= "Rep protein"
 124..125
 /tag= c
 /note= "Terminal resolve site (TRS)"
 219..226
 /*tag= d
 /bound moiety= "USF"
 /note= "E box"
 236..239
 /*tag= e
 /label= P5 promoter
 237..245
 /*tag= f
 /bound moiety= "YY1 factor"
 270..275
 /tag= g
 /label= P5_TATA-Box
 299..306
 /*tag= h
 /note= "YY1/p5 RNA"
 335..2206
 /*tag= i
 /product= "Rep 78"
 /function= "regulates replication and integration of AAV DNA into host cell's chromosome"
 335..2272
 /*tag= j
 /product= "Rep 68"
 /function= "regulates replication and integration of AAV DNA into host cell's chromosome"
 /note= "The coding region is interrupted by intron"
 1924..2220
 /*tag= k
 /note= "This region interrupts the coding sequence of Rep 68 and Rep 40"
 857..862
 /*tag= l
 /label= P19_TATA_Box
 882..883
 /*tag= m
 /note= "P19 RNA"
 1007..2206
 /*tag= n
 /product= "Rep 52"
 /function= "regulates replication and integration of AAV DNA into host cell's chromosome"
 1007..2272
 /*tag= o
 /product= "Rep 40"
 /function= "regulates replication and integration of AAV DNA into host cell's chromosome"
 /note= "The coding region is interrupted by intron"
 1836..1841
 /*tag= p
 /label= P40_TATA-BOX
 1875..1876
 /*tag= q
 /note= "P40 RNA"
 2223..4433
 /*tag= r
 /product= "VP1 protein"
 /note= "Capsid protein"
 2634..4433
 /*tag= s
 /product= "VP2 protein"
 /note= "Capsid protein"
 2829..4433
 /*tag= t


```

FT /product= "VP3 protein"
FT /note= "Capsid protein"
FT 4447...4452
FT /tag= u
FT 4576...4718
FT /tag= v
FT /label= 3' ITR
FT /note= "Inverted terminal repeat which is capable
FT of forming T-shaped hairpin structure"
XX
PN WO200028061-A2.
XX
XX 18-MAY-2000.
XX
XX 02-NOV-1999; 99WO-US25694.
XX
XX 05-NOV-1998; 98US-0107114.
XX
XX (UYPE-) UNIV PENNSYLVANIA.
XX
XX Wilson JM, Xiao W;
PI
XX
XX WPI; 2000-376571/32.
DR P-PSDB; AAY71161, AAY71164, AAY71165, AAY71166, AAY71167, AAY71168,
DR AAY71169.
XX
XX Novel adeno-associated virus serotype 1 polynucleotide useful for
PT preparation of medicament for delivery of a transgene to a host -
XX
XX Claim 1; Fig 1; 108pp; English.
XX
XX The present sequence is an adeno-associated virus serotype 1 (AAV-1)
CC DNA characterised by two inverted terminal repeats (ITR) and open
CC reading frames for rep and capsid (cap) proteins. The rep reading frame
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.
CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap
CC coding regions, are useful in production of recombinant viral vectors
CC for gene delivery. These vectors can be used as gene therapy
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
CC not induce the formation of neutralising antibodies specific to any
CC serotype of AAV hence is useful for transforming host cells, and in the
CC preparation of a medicament for the delivery of transgene to a host.
XX
XX Sequence 4718 BP; 1121 A; 1393 C; 1273 G; 931 T; 0 other;
SQ

```

Alignment Scores:

| Pred. No.: | 6.23e-283 | Length: | 4718 |
|------------------------|-----------|---------------|------|
| Score: | 3989.00 | Matches: | 736 |
| Percent Similarity: | 100.00% | Conservative: | 0 |
| Best Local Similarity: | 100.00% | Mismatches: | 0 |
| Query Match: | 100.00% | Indels: | 0 |
| DB: | 21 | Gaps: | 0 |

US-09-807-802A-13 (1-736) x AAD00772 (1-4718)

```

QY 1 MetAlaAlaAspGlyTyrLeuProAspTyrLeuGluAspAsnLeuSerGluGlyIleArg 20
DB 2223 ATGGCTGCCGATGGTATCTTCAGATTGGCTCGAGGACAACTCTCTGAGGGGCAATCGC 2282
QY 21 GluTyrTyrAspLeuAspProGlyAlaProLysPheAlaAsnGlnGlnLysGlnAsp 40
DB 2283 GAGTGTGGGACTTGAACCTGGAGCCCGAGGCCCAAGCCAAAGCCAGCAAGCAGGAC 2342
QY 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
DB 2343 GACGGCGGGGCTGGTGTCTTCAGTACCTCGACCCCTTCAACGGGCTCGAC 2402
QY 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
DB 2403 AAGGGGAGCCCCGTCAACGGCGGAGCGAGCGGCCCTTGAGCAACGACGAGGCTACGAC 2462
QY 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100

```

```

DB 2463 CAGCAGCTCAAGCGGGTGACAATCCGTAACCTCGGTATATACCAACCGCGAGTGT 2522
QY 101 GlnGluArgLeuGlnGluAspThrSerPheGlyAsnLeuGlyArgAlaValPheGln 120
DB 2523 CAGGAGCGTCTGCAAGAGATACGTCTTTTGGGGGCAACCTCGGGCGAGAGTCTTCCAG 2582
QY 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaPro 140
DB 2583 GCCAAGAAAGCGGGTCTCGAACCTCTCGGTCTGGTTGAGGAGGGCGCTAAGACGGCTCT 2642
QY 141 GlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSerGlyIleGly 160
DB 2643 GGAAGAAACGTCGGTAGAGCAGTCGCCACACAGAGCCAGACTCTCTCTCGGGCATCGGC 2702
QY 161 LysThrGlyGlnGlnProAlaLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
DB 2703 AAGACAGCCAGCAGCCCGCTAAAGAGAGACTCAATTTTGGTTCAGACTGGGCACTCAG 2762
QY 181 SerValProAspProGlnProLeuGlyGluProAlaThrProAlaAlaValGlyPro 200
DB 2763 TCAGTCCCGATCCACAACCTCTCGAGAACTCCAGCAACCCCGCTGCTGTGGACCT 2822
QY 201 ThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGlnGlyAlaAspGly 220
DB 2823 ACTCAATGGCTTCAGGCGGTGGCGCACCAATGGCAGACAATAACGAAGGCGCGCAGCGA 2882
QY 221 ValGlyAsnAlaSerGlyAsnTyrHisCysAspSerThrTyrLeuGlyAspArgValIle 240
DB 2883 GTGGGTAAATGGCTCAGGAAATTTGGCATTTGCCATTTCCACATGGCTGGGCGAGAGTCA 2942
QY 241 ThrThrSerThrArgThrTyrAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIle 260
DB 2943 ACCACGACAGCCCGCACTGGCGCTTGGCCACCTTACCAATAACCACTCTACAGCAATC 3002
QY 261 SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTyr 280
DB 3003 TCCAGTGTCTCAACGGGGGCGCAGCAACGACCACTACTTTCGGCTACAGCACCCCTGG 3062
QY 281 GlyTyrPheAspPheAsnAspPheHisCysHisPheSerProArgAspTyrGlnArgLeu 300
DB 3063 GGGTATTTGATTTCAACAGATTTCCAGTTCACCTTTTCCACCGTGTGATCGCAGGACTC 3122
QY 301 IleAsnAsnAsnTyrGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGln 320
DB 3123 ATCAACAACATTTGGGATTCGGCCCAAGAGACTCAACTTCAACTCTTCAACATCAA 3182
QY 321 ValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThr 340
DB 3183 GTCAAGGAGGTCACGACGAAATGATGGCGTCAACACCATCGCTTAATACCTTACCGAGC 3242
QY 341 ValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGln 360
DB 3243 GTTCAAGTCTTCTCGACTCGGACTACGAGTTCGGTACGTCTCGGTCTGGGACCCAG 3302
QY 361 GlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr 380
DB 3303 GGTCGCTCTCTCGCTTCCGGCGGACGTTTCATGATTCGCAATACGGCTACCTGAGC 3362
QY 381 LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 400
DB 3363 CTCACAATGGCAGCAACCGCTGGGAGTTCATCTCTTTTACTGCTCGAATATTTCCT 3422
QY 401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 420
DB 3423 TCTCAGATCTGAGAACGGGCAACAACTTTACCTTCAGCTACCTTTGAGGAAGTGCCT 3482
QY 421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
DB 3483 TTCCACAGCAGCTACCGCACAGCCAGAGCTGGACCGGCTGATGAATCTCTCTACGAC 3542
QY 441 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAsp 460

```

Db 3543 CAATACCTCTATTACCTGAACAGAACTCAAAATCAGTCGGAGTGGCCCAAAACAAAGGAC 3602
 Qy LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuPro 480
 Db 3603 TTGCTGTTTAGCGGTGGTCTCCAGCTGCGATGCTGTCTTTCAGCCCAAAACTGGCTACCT 3662
 Qy GlyProCysGlyArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsn 500
 Db 3663 GGACCCCTGTTATCGGCAGCAGCGCGTTTCTTAACAAACAAACAGACAAACACACAGCAAT 3722
 Qy PheThrTrpThrGlyValaSerLysValAsnLeuAsnGlyArgGluSerIleIleAsnPro 520
 Db 3723 TTACTCGACTGCTGCTTCAAAATATAACCTCAATGGCGGTGAATCCATCATCAACCT 3782
 Qy GlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyVal 540
 Db 3783 GGCACCTGCTATGSCCTCACAAAGACGACGAACAGTCTTCTTCCATGAGCGGTGC 3842
 Qy MetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIle 560
 Db 3843 ATGATTTTGGAAAGAGAGCGCGGAGCTTCAACACACTGCATTGGACAAATGTCATGATT 3902
 Qy ThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrVal 580
 Db 3903 ACAGACGAGAGGAAATTAAGCCACTTAACCTGTGGCCACCGAAGATTTGGACCGTG 3962
 Qy AlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGly 600
 Db 3963 GCAGTCAATTTCCAGAGCAGCAGACAGACCCCTGCGACCGAGATGTCATGTGGA 4022
 Qy AlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAla 620
 Db 4023 GCATTACCTGGCATGTGTGGCAAGATAGAGACGTGTACTGCGAGGTCCTCATTTGGGCG 4082
 Qy LysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeu 640
 Db 4083 AAAATTCCTCACACAGATGACACTTTCACCCGCTCTCTTATGGCGCGCTTGGACTC 4142
 Qy LysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAla 660
 Db 4143 AAGAACCCGCTCTCTCAGATCTCTCATCAAAACACAGCCTGTCTCGAATCTCCGGCG 4202
 Qy GluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSer 680
 Db 4203 GAGTTTTCAGCTACAAAGTTTGCTTATTTCATCACCCTTACTCCACAGACAAGTGAGT 4262
 Qy ValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGln 700
 Db 4263 GTGGAAATTTGAATGGGAGCTGCGAAGAAAGAAACAGACGCTGGATCCCGAAGTGCAG 4322
 Qy TyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnGlyLeu 720
 Db 4323 TACACATCCAAATATGCAAAATCTGCCAAGTTGATTTTACTGTGGACAAACATGAGATT 4382
 Qy TyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
 Db 4383 TATACTGAGCTCGCCCCATTGGACCCGTTACTTACCCTGCCCTG 4430

RESULT 3

AAF23749

ID AAF23749 standard; DNA; 4683 BP.

XX AC

XX AAF23749;

XX AC

DT 28-MAR-2001 (first entry)

XX XX

XX DE

XX AAV6 DNA sequence.

XX XX

AAV6; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
 atherosclerosis; sickle cell anemia; thalassemia;
 blood clotting disorder; diabetes; ss.

XX OS

Adeno associated virus.

XX

PN US6156303-A.

XX

PD 05-DEC-2000.

XX

PF 11-JUN-1997; 97US-0873168.

XX

PR 11-JUN-1997; 97US-0873168.

XX

PA (UNIW) UNIV WASHINGTON.

XX

PI Russell DW, Rutledge EA;

XX

WPI; 2001-060164/07.

DR

Adeno-associated virus serotype 6 and viral vector derived from it for
 gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
 syndrome, sickle cell anemia, thalassemia and diabetes

XX

Claim 1; Fig 1; 50pp; English.

CC

The present invention relates to adeno-associated virus serotypes. The
 present sequence is the DNA sequence of one such serotype (AAV6). AAV6
 can be used to construct AAV viral vectors for use in gene therapy for a
 range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
 sickle cell anemia, thalassemia, blood clotting disorders and diabetes.
 The AAV viral vectors have increased transduction efficiency of a
 particular host cell as the AAV virion containing the AAV vector genome
 can be modified to express a capsid protein of an AAV serotype that
 transduces the selected host cell.

SQ

Sequence 4683 BP; 1114 A; 1363 C; 1277 G; 929 T; 0 other;

Alignment Scores:

Pred. No.: 5e-281 Length: 4683
 Score: 3963.00 Matches: 730
 Percent Similarity: 99.59% Conservative: 3
 Best Local Similarity: 99.18% Mismatches: 3
 Query Match: 99.35% Indels: 0
 DB: 22 Gaps: 0

US-09-807-802A-13 (1-736) x AAF23749 (1-4683)

Qy

1 MetAlaAlaAspGlyTyrLeuProAspTrpLeuGluAspAsnLeuSerGlyIleArg 20

Db

2208 ATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGGACAACCTCTCTGAGGGCATTCGC 2267

Qy

21 GluTrpTrpAspLeuLysProGlyAlaProLysProLysAlaAsnGlnGlnLysGlnArg 40

Db

2268 GAGTGGTGGGACTTGAACTGGAGCCCGGAAACCCAAAGCCACACGCAAAAGCAGGAC 2327

Qy

41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuArg 60

Db

2328 GACGCCGGGGTCTGGTCTTCTCGCTACAGTAGTACCTCGGACCTTCAACGAGCTCGAC 2387

Qy

61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrArg 80

Db

2388 AAGGGGGAGCCCGTCAACGCGCGGATGACGGGGCCCTCGAGCAGCAGCAAGCCCTACGAC 2447

Qy

81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100

Db

2448 CAGCAGCTCAAGCGGGTGACAAATCCGTACCTCGGGTATACCAACGCCGCGCGAGTTT 2507

Qy

101 GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln 120

Db

2508 CAGGAGCGTCTGCAAGAAGATACGTCCTTTGGGGGCAACCTCGGGCGAGCAGCTCTTCCAG 2567

Qy

121 AlaLysArgValLeuGluProLeuGlyLeuValGluGlyValAlaLysThrAlaPro 140

Db

2568 GCCAAGAAAGAGGTTCTCGAACCTTTTGGTCTGGTTGAGAGGGTCTTAAGACGGCTCTCT 2627

Qy

141 GlyLysArgProValGluGlnSerProGlnGluProAspSerSerGlyIleGly 160

2628 GGAAGAAACGTCGGTAGACAGTCGCCACAAGACCGACAGCTCCTCTCGGGCATTGGC 2687
161 LysThrGlyGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
2688 AAGACAGGCGCAGCGCCCTAAAAGAGACTCAATTTTGGTTCAGACTGGCGACTCAGAG 2747
181 SerValProAspProGlnProLeuGlyGluProAlaThrProAlaAlaValGlyPro 200
2748 TCAGTCCCCCGACCAACACTCTCGGAGAACTCCAGCAACCCCGCTGCTGTGGGACT 2807
201 ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGly 220
2808 ACTACAATGGCTTCAGCGGTGGCGCACCAATGGCAGACAATAACGAAGCGCCGACGGA 2867
221 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIle 240
2868 GTGGGTAAATGGCTTCAGGAATTTGGCATTTCCATGCTGGGCGACAGATCATC 2927
241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIle 260
2928 ACCACGACGACCGCAACATGGGCTTGGCCCACTATACCAACCACTCTACAGCAATC 2987
261 SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrp 280
2988 TCCAGTGTCTCAACGGGGCGCAGCAACGACCACTACTCTGGCTTACAGCAACCCCTGG 3047
281 GlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu 300
3048 GGGTATTTTGAATTTCAACAGATTTCCACTGCCATTTCTCCACGCTGACGCGACGCTC 3107
301 IleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGln 320
3108 ATCAACACCAATTTGGGGATTCGGCCCAAGAGACTCACTTCAAGCTCTTCAACATCCA 3167
321 ValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThr 340
3168 GTCAAGGAGTCAACGCAATGATGGGTGCGTCACGACCATGCTTAATACCTTACGACGCG 3227
341 ValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGln 360
3228 GTTCAAGTCTTCTCGGACTCGGAGTACCACTGTCCTGCTCGGCTCGGCGACCGAG 3287
361 GlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr 380
3288 GGCTGCTCTCTCGCTTCCGGCGGAGCGTTCATGATTCGGCAGTACGCTACCTAACG 3347
381 LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 400
3348 CTCACAAATGGCAGCCAGCGCAGTGGGACGGTCACTCTTTACTGCTCGCTGGAATATTTCCA 3407
401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 420
3408 TCGCAGATGCTGGAAGCGGCAATACTTTTACCTTTCAGCTTACCTTCGAGGACGTGCT 3467
421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
3468 TTCCACGACGACTACCGGCACAGCCAGCGCTGGACCGGCTGATGATCTCTCATCGAC 3527
441 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAsp 460
3528 CAGTACCTGTATTACTTGAACAGAACTCAGAACTCAGTCCGGAAGTCCCAAAACAAGGAC 3587
461 LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuPro 480
3588 TTGCTGTTTAGCCGGGGGTCTCCAGCTGGCATGTCTGTTTCCAGCCCAAACTGGCTACCT 3647
481 GlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsn 500
3648 GGACCTGTGTACCGGCAGCGCGCTTTCTTAAACAAAAACAGCAACCAACGCAAC 3707
501 PheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnPro 520
3708 TTTACCTGGAGCTGGTGTCTCAAAATATAACCTTAAATGGGCGTGAATCTATAATCAACCT 3767

QY 521 GlyThrAlaMetAlaSerHisLysAspGluAspLysPhePheProMetSerGlyVal 540
DB 3768 GGCACCTGCTATGGCTCACCAAGACGACAAAGACAAGTCTCTTCCATGAGCGGTGC 3827
QY 541 MetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIle 560
DB 3828 ATGATTTTGGAAAGGAGAGCGCGGAGCTTCAAAACACTGCATTTGGACAATGTCTATGC 3887
QY 561 ThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrVal 580
DB 3888 ACAGACGAGAGGAAATCAAGCCACTAACCCCTGGCCACCCGAAAGATTTGGGACCTGTG 3947
QY 581 AlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGly 600
DB 3948 GCATGCAATCTCCAGAGCAGCAGCAGCCTCGACCGGAGATGTGCATGTATGGGA 4007
QY 601 AlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAla 620
DB 4008 GCCTTACTGGAATGGTGTGGCAAGACAGAGACGTATACCTGCAGGGTCTTATTTGGGCC 4067
QY 621 LysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeu 640
DB 4068 AAAATTTCTCACGCGATGGACACTTTCACCCGCTCTCTCATGGGCGGCTTTGGACTT 4127
QY 641 LysAsnProProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAla 660
DB 4128 AAGCACCCTCTCTCAGATCTCTCATCAAAACACGCTGTCTCTCGGAATCTCTCCGCA 4187
QY 661 GluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSer 680
DB 4188 GAGTTTTCGGCTACAAAGTTTGTCTTATTCATCACCAGTATTCACAGGACAAAGTGAGC 4247
QY 681 ValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGln 700
DB 4248 GTGGAGATTGAATGGGAGCTGCAGAAAGAAACAGCAACGCTGGAATCCCGAAGTGCGAG 4307
QY 701 TyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeu 720
DB 4308 TATACATCTACTATGCAAAATCTGCCAATCTTGATTTCACTGTGGACAACAATGGACTT 4367
QY 721 TyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
DB 4368 TATACTGAGCTCGCCCATTTGGCACCCTTACCTCACCCGCTCCCTG 4415

RESULT 4

AAF23748
ID AAF23748 standard; DNA; 4722 BP.

AC AAF23748;

DT 28-MAR-2001 (first entry)

XX AAV3B DNA sequence.

XX AAV3B; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
KW atherosclerosis; sickle cell anaemia; thalassaemia;
KW blood clotting disorder; diabetes; ss.

XX Adeno associated virus.

XX US6156303-A.

XX 05-DEC-2000.

XX 11-JUN-1997; 97US-0873168.

XX 11-JUN-1997; 97US-0873168.

XX (UNIW) UNIV WASHINGTON.

XX Russell DW, Rutledge EA;

XX

DR WPI: 2001-060164/07.

Adeno-associated virus serotype 6 and viral vector derived from it for gene therapy of cystic fibrosis, cancer, acquired immunodeficiency syndrome, sickle cell anemia, thalassemia and diabetes -

PS Example 2; Fig 1; 50pp; English.

The present invention relates to adeno-associated virus serotypes. The present sequence is the DNA sequence of one such serotype (AAV3B) can be used to construct AAV viral vectors for use in gene therapy for a range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis, sickle cell anaemia, thalassemia, blood clotting disorders and diabetes. The AAV viral vectors have increased transduction efficiency of a particular host cell as the AAV virion containing the AAV vector genome can be modified to express a capsid protein of an AAV serotype that transduces the selected host cell.

Sequence 4722 BP; 1204 A; 1297 C; 1243 G; 978 T; 0 other;

Alignment Scores:

| | | | |
|------------------------|-----------|---------------|------|
| Pred. No.: | 7,766-248 | Length: | 4722 |
| Score: | 3511.00 | Matches: | 640 |
| Percent Similarity: | 91.99% | Conservative: | 38 |
| Best Local Similarity: | 86.84% | Mismatches: | 57 |
| Query Match: | 88.02% | Indels: | 2 |
| DB: | 22 | Gaps: | 2 |

US-09-807-802A-13 (1-736) x AAF23748 (1-4722)

| | | | |
|----|------|--|------|
| Qy | 1 | MetAlaAlaAspGlyTyrLeuProAspTrpLeuGluAspAsnLeuSerGluGlyIleArg | 20 |
| Db | 2208 | ATGGCTGCTCAGCGTTATCTTCCAGATTGGCTCGAGGACAACTTTCTGAAGGCAATTCGT | 2267 |
| Qy | 21 | GluTrpTrpAspLeuLysProGlyAlaProLysProLysAlaAsnGlnGlnLysGlnAsp | 40 |
| Db | 2268 | GAGTGGTGGGCTCTGAACCTGGAGTCCCTCAACCCAAAGCGAACCAACACACCGACGAC | 2327 |
| Qy | 41 | AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp | 60 |
| Db | 2328 | AACCGTCGGGGTCTTGTTGCTTCGGGTTCACAAATACCTCGGACCGGTAAACGGACTCGAC | 2387 |
| Qy | 61 | LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaIleTyrAsp | 80 |
| Db | 2388 | AAAGGAGAGCGGTCAACGAGGGCGGACGGCGACCCCTCGAACACGCAAAAGCTTTACGAC | 2447 |
| Qy | 81 | GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe | 100 |
| Db | 2448 | CAGCAGCTCAAGCGCGGTGACNACCCCTCAAGTACNACCACGCGCGCGGAGTTT | 2507 |
| Qy | 101 | GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln | 120 |
| Db | 2508 | CAGGAGCGCTCTTCAAGAAGATACGCTTTTGGGGGCAACCTTGGCAGAGCAGTCTTCCAG | 2567 |
| Qy | 121 | AlaLysLysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaPro | 140 |
| Db | 2568 | GCCAAAGAAGAGATCTTGAGCCTCTTGGTCTGGTTCGAGGAGCAGCTAAACGGCTCCT | 2627 |
| Qy | 141 | GlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSerGlyIleGly | 160 |
| Db | 2628 | GGAAAGAAGAGGCCCTGTAGATCAGTCTCCTCAGGAACCGGACTCATCATCTGGTGTGGC | 2687 |
| Qy | 161 | LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu | 180 |
| Db | 2688 | AAATCGGGGCAACAGCTGCCAGAAAAGAGACTAAATTTTCGGTTCAGACTGGCGACTCAGAG | 2747 |
| Qy | 181 | SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro | 200 |
| Db | 2748 | TCAGTCCAGNACCCCTCAACCTCTCGGAGAACCAACAGCAGGACCCCAACAGTTTGGGATCT | 2807 |
| Qy | 201 | ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGly | 220 |
| Db | 2808 | AAATCAATGGCTTCAGCGGTGGCGCAACCAATGTCAGACAATAACAGGGGTGCCGATGGA | 2867 |

Query Match: 86.21% Indels: 2
DB: 21 Gaps: 2
US-09-807-802A-13 (1-736) x AAD00834 (1-8178)

QY 1 MetAlaAlaAspGlyTyrLeuProAspTrpLeuGluAspAsnLeuSerGluGlyIleArg 20
DB 2133 ATGGCTGCCGATGGTGTATCTTCAGATTGGCTCGAGGACACTCTCTCTGAAGGAATAAGA 2192

QY 21 GluTrpTrpAspLeuysProGlyAlaProLysPhePheAlaAsnGlnGlnLysGlnAsp 40
DB 2193 CAGTGGTGGAGCTCAAACTGGCCACCACCAACGAGCCCGCAGAGCGGATAGGAC 2252

QY 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
DB 2253 GACAGCAGGGGCTTCTGGTCTCTCGGTACAAGTACCTCGGACCTTCAACGAGCTCGAC 2312

QY 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
DB 2313 AAGGGAGAGCCGGTCAACGAGGAGCGCGCGGCCCTCGAGCAGCACAAGCCTACGAC 2372

QY 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
DB 2373 CGGCACCTCGACAGCGGAGCAACCCGTACCTCAAGTACAACACCGCCGCGGGGTTT 2432

QY 101 GlnGluArgGlnGlnAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln 120
DB 2433 CAGGAGCGGCTTAAAGAAAGATACGCTCTTTGGGGGCAACCTCGGACGAGCAGCTCTCCAG 2492

QY 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaPro 140
DB 2493 GCGAAAGAGGGGTTCTTGAACCTCTGGGCTGGTTGAGGAACCTGTTAAGACGGCTCGG 2552

QY 141 GlyLysLysArgProValGluGlnSerProGlnProAspSerSerGlyIleGly 160
DB 2553 GGAAGAAAGAGGCGGTAGACACTCTCTGTGGACCCAGACTCTCTCTCGGGAACCGGA 2612

QY 161 LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
DB 2613 AAGCGGGCGCAGCAGCTCGCAAGAAAGATTGAATTTTGGTCAGACTGGAGACGCGAGC 2672

QY 181 SerValProAspProGlnProLeuGlyGluProAlaThrProAlaThrProAlaValGlyPro 200
DB 2673 TCAGTACCTGACCCCGACCTCTCGGACAGCCACCGACGCCCCCTCTCGTCTGGGAAC 2732

QY 201 ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnGluGlyAlaAspGly 220
DB 2733 AATACGATGGCTACAGCAGTGGCGCACCAATGGCAGACAATAACGAGGCGCGCGGGA 2792

QY 221 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIle 240
DB 2793 GTGGGTAAATTCCTCCGGAAATTTGGCATTTGGATTCACATGGATGGGCGCAGAGTCA 2852

QY 241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIle 260
DB 2853 ACCACGACGACCCGAACTGGGCGCTGGCCCTCCACCTCAACACCAACCTCTACAAACAA 2912

QY 261 SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrp 280
DB 2913 TCCAGCCAAATCA--GGAGCCTCGAACGACAATCACTACTTTGGCTACAGCACCCTTGG 2969

QY 281 GlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu 300
DB 2970 GGGTATTTTGGACTTCAACAGATTCCACTGGCACTTTTCCACCTACCGAGCTGCGAAAG 3029

QY 301 IleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGln 320
DB 3030 ATCACACACAACTGGGGATTTCCGACCCCAAGAGACTCAACTCAAGCTCTTTACATTCAA 3089

QY 321 ValLysGluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeuThrSerThr 340
DB 3090 GTCAAGAGGTCACGCGAATGACGGTACGACGACGATTTGCCAATAACCTTACCAGCAG 3149

QY 341 ValGlnValPheSerAspSerGlyTyrGlnLeuProTyrValLeuGlySerAlaHisGln 360
DB 3150 GTTCAGGTGTTTACTGACTCGAGTACCGAGTCCGCTACGTCTCGGCTCGGCGCACCAA 3209

QY 361 GlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr 380
DB 3210 GGCTGTCTCCCGCGTTCAGCGGACGCTCTTCATGGTCCCTCAGTATGGATACCTCACC 3269

QY 381 LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 400
DB 3270 CTGAACACGGAAGTCAAGCGGTGGAGCTCATCTTTTACTGCTGGAGTACTTCCT 3329

QY 401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 420
DB 3330 TCGCAGATGCTAAGGACTGGAAATAACTTCCAATTCAGCTATACCTTCGAGGATGTACT 3389

QY 421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
DB 3390 TTTTACAGCAGCTACGCTCACAGCAGAGTTTGGATCGCTTGTATGATTCCTCTTATGAT 3449

QY 441 GlnTyrLeuTyrTyrLeuAsnArgThrGln--AsnGlnSerGlySerAlaGlnAsnLys 459
DB 3450 CAGTATCTGTACTACCTGGAACAGAGCGCAAGCAACCTCTTGGAAACAACCAACATCA 3509

QY 460 AspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeu 479
DB 3510 CGGCTGCTTTTATAGCGAGCTGGGCTCAGTCTATGCTTTTGGAGCCCAAGAAATTTGGT 3569

QY 480 ProGlyProCysTyrArgGlnArgValSerLysThrLysThrAspAsnAsnAsnSer 499
DB 3570 CTTGGGCTCTGCTACCGCAACAGAGACTTTCAAGAGCTGCTAACGACACACAAACAGT 3629

QY 500 AsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsn 519
DB 3630 AACTTTCTTGACAGCGGCCAGCAAAATATCATCTCAATGCGCGGACTCGTGGTGAAT 3689

QY 520 ProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGly 539
DB 3690 CCAGGACAGCTATGGCCAGTCACAAGAGCATGAAGAAATAATTTTCCCTATGACCGGC 3749

QY 540 ValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMet 559
DB 3750 AATCTAATTTTGGCAAGAGGACACGCGAGTAAACGAGATTAAGATTAATGTAATG 3809

QY 560 IleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThr 579
DB 3810 ATTACGATGAAGAAGAGATTCTGACCAACCAATCTCTGTGGCAACAGAGCAGTATGAACT 3869

QY 580 ValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMet 599
DB 3870 GTGGCAAAATACTTGCAGAGCTCAAAATACAGCTCCACGACTGGAACTGTCAATCATCAG 3929

QY 600 GlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrp 619
DB 3930 GGGGCTTACTCTGGCATGGTGTGGCAAGATCTGTACGTGTACTTCAAGGACCTATCTGG 3989

QY 620 AlaTyrIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGly 639
DB 3990 GCAAAATTCCTCACACGAGTGGACACTTTCATCTCTCTCTCTGATGGAGGCTTGA 4049

QY 640 LeuLysAsnProProProGlnIleLeuIleLysAsnThrProValProAlaAsnProPro 659
DB 4050 CTGAACATCCGCTCTCTCAATCATGATCAAAATACTCCGGTACCTGCGGAATCTTCG 4109

QY 660 AlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnVal 679
DB 4110 ACCACTTTCAGTGGGCAAGATTGCTTCTTCTCATCACAGTACTTCCACGGGACAGGTC 4169

QY 680 SerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluVal 699
DB 4170 AGCTGGAGATCGAGTGGAGCTGCAGAAAGAAACAGCAAAACGCTGGAATCCCGAAAT 4229

QY 700 GlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGly 719

```

Db 4230 CAGTACACTTCCAACTACCAACAGTCTGTTAATCGTGGACTTACCGTGGATTAATGGC 4289
Qy 720 LeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 4290 GTGATTTCAGAGCTCGCCCAATGGCCACCATGCTGACTCGTAATCTG 4340

RESULT 6
ID AAH26327
ID AAH26327 standard; DNA; 4072 BP.
XX
AC AAH26327;
XX
DT 02-OCT-2001 (first entry)
XX
DE Adenovirus helper Ad cap2.
XX
KW AAV; vector; adenovirus; helper virus; Ad cap2; loxP site;
KW Gene therapy; ds.
XX
OS Chimeric - Mastadenovirus.
OS Chimeric - Adeno associated virus.
OS Chimeric - Human cytomegalovirus.
XX
PN W0200155361-A2.
XX
PD 02-AUG-2001.
XX
PF 26-JAN-2001; 2001WO-US02709.
XX
PR 26-JAN-2000; 2000US-0178536.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Hardy SE;
XX
DR WPI; 2001-483239/52.
XX
PT Producing recombinant adeno-associated virus (rAAV) vector, by stably
PT infecting eukaryotic host cell with rAAV vectors, helper adenovirus or
PT herpes virus, recombinant adenovirus or herpes vectors -
XX
Example 2; Page 58-59; 63pp; English.
XX
CC The present sequence is that of adenovirus helper Ad cap2, made
CC by Cre/lox recombination and expressing adeno-associated virus
CC (AAV) VP1, 2,3 from the cytomegalovirus immediate early promoter
CC of pAdlox, and containing a loxP site. Ad cap2 was used to
CC demonstrate recombinant AAV (rAAV) production from virally
CC transduced cells. The invention provides methods and compositions
CC for producing rAAV vector particles by: (a) introducing into a host
CC cell (i) AAV packaging plasmid pFloxAAV (see AAH26326), (ii) a
CC recombinant viral vector encoding plasmid, and (iii) a plasmid
CC encoding herpesvirus, cytomegalovirus or adenoviral functions, or
CC a herpesvirus, cytomegalovirus or adenovirus itself, to produce
CC a flox AAV particles and rAAV particles; and (b) introducing into a
CC second host cell (i) the rAAV particles or (a), (ii) a vector
CC that directs expression of Cre, and (ii) a vector which directs
CC expression of herpes virus, cytomegalovirus or adenovirus helper
CC functions, such that rAAV vector particles are produced. The
CC vectors are useful for in vivo or in vitro gene therapy and also
CC for in vitro recombinant protein production.
XX
SQ Sequence 4072 BP; 1073 A; 1025 C; 992 G; 982 T; 0 other;

Alignment Scores:
Pred. No.: 5.95e-240 Length: 4072
Score: 3402.50 Matches: 613
Percent Similarity: 90.22% Conservative: 51
Best Local Similarity: 83.29% Mismatches: 71
Query Match: 85.30% Indels: 1
DB: 22 Gaps: 1

```

```

US-09-807-802A-13 (1-736) x AAH26327 (1-4072)
Qy 1 MetAlaAlaAspGlyTyrLeuProAspTyrLeuGluAspAsnLeuSerGluGlyLeuArg 20
Db 1484 ATGCTCGCGATGGTTATCTTCCAGATGGCTCGAGGACACTCTCTCTGAAGGAATAGA 1543
Qy 21 GluTyrTrpAspLeuLeuProGlyValAlaProLysProLysAlaAsnGlnGlnLeuAsp 40
Db 1544 CAGTGGTGGAGCTCAAAACCTGGCCCAACCAACCCGAGAGCGGCGGCGGAGGAC 1603
Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
Db 1604 GACAGCAGGGGTCTTGTGCTTCTGGGTACAGTACCTCGGACCTTCAACGGACTCGAC 1663
Qy 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
Db 1664 AAGGAGAGCCGGTCAACGAGGAGAGCGCCGGCCCTCGAGACGACCAAGCCTACGAC 1723
Qy 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
Db 1724 CGGCAGCTCGACAGCGGAGACCAACCCGTACCTCAAGTACCAACCAACCGCGCGAGTTT 1783
Qy 101 GlnGluArgLeuGlnGluAspThrSerPheGlyAsnLeuGlyArgAlaValPheGln 120
Db 1784 CAGGAGCGCTTAAAGAGATACGTCTTTTGGGGGCAACCTCGGACGAGCAGTCTTCCAG 1843
Qy 121 AlaLysLysArgValLeuGluProLeuGlyValGluGluGlyAlaLysThrAlaPro 140
Db 1844 CGGAAAAGAGGGTCTTGAACCTCTGGGCTCTGGTGGAGAACCTGTGTAAGCGGCTCG 1903
Qy 141 GlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSerGlyLeuGly 160
Db 1904 GGAAAAAAGAGCGGTAGACACTCTCTCTGTGAGCCAGACTCTCTCTCGGAAACCGGA 1963
Qy 161 LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
Db 1964 AAGCGGGCCAGCAGCCTGCAAGAAAAGATTGAATTTTGGTTCAGACTGGAGACCGCAG 2023
Qy 181 SerValProAspProGlnProLeuGlyGluProAlaThrProAlaAlaValGlyPro 200
Db 2024 TCAGTACTGTGACCCCGACCTCTCGGACAGCCACCCAGCAGCCCTCTGGTGGGAATC 2083
Qy 201 ThrThrMetAlaSerGlyGlyValAlaProMetAlaAspAsnAsnGlyGlyAlaAspGly 220
Db 2084 AATACGATGGCTACAGCGCAGTGGCGCACAATGGCAGACAATAACGAGGCGCGCCGCGA 2143
Qy 221 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIle 240
Db 2144 GTGGGTAAATTCCTCGGAAATTTGGCATTTGGCATTTCCATGATGGGCGACAGATCATC 2203
Qy 241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIle 260
Db 2204 ACCACGACGACCGGAACCTGGGCCCTGGCCCTACCAACCAACCACTCTTACAACAATTT 2263
Qy 261 SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrp 280
Db 2264 TCCAGCCCAATCA--GGAGCCTCGAACGACAATCACTATTTTGGCTACAGCACCCTTGG 2320
Qy 281 GlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu 300
Db 2321 GGGTATTTTGACTTCAACAGATTCCACTGCCACTTTTCCACGACGATGCTGGCAAGACTC 2380
Qy 301 IleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGln 320
Db 2381 ATCAACACAACTGGGGATTCCGACCCCAAGAGACTCAACTTCAAGCTCTTTTAACATCAA 2440
Qy 321 ValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThr 340
Db 2441 GTCAAGAGGTCACGCAGAATGACGGTACGACGACGATTTGCCAATAAACCCTTACGACAG 2500
Qy 341 ValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGln 360
Db 2501 GTTTCAGGTGTTTACTCGGAGTACGAGCTCCCGTACGCTCTCGGCTCGGCGGATCAA 2560

```



```
QY 361 GlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr 380
Db 2561 GGATGCTCCCGCCGCTCCAGCAGACGCTTCATGGTGCCACAGTATGATACCTCACC 2620
QY 381 LeuAsnAnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 400
Db 2621 CTGAACAACGGGAGTCAGGACAGTAGGACGCTCTTCATTTACTGCTGGAGTACTTTCCT 2680
QY 401 SerGlnMetLeuArgThrGlyAsnAnPheThrPheSerTyrThrPheGluGluValPro 420
Db 2681 TCTCAGATGCTCGTACCGGAACACTTTACCTTCAGCTACACTTTTGGAGACGCTTCT 2740
QY 421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
Db 2741 TTCCACAGCAGCTACGCTCACAGCCAGAGTCTGGACCGCTCTCATGAATCCTCTCATCGAC 2800
QY 441 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAnGlnSerGlySerAlaGlnAsnLysAsp 460
Db 2801 CAGTACCTGTATTACTTGAGCAGAAACAACACTCCCAAGTGGAAACCCACGCGCTCAAGG 2860
QY 461 LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuPro 480
Db 2861 CTTCAGTTTCTCAGCCGCGAGCTACCAAGTACCACTTCGAGCAGCTCTAGGAACCTGGCTTCT 2920
QY 481 GlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsn 500
Db 2921 GGACCTCTGTACCGCAGCAGCGAGTATCAAGACATCTCGCGATTAACAACAACAGTGAA 2980
QY 501 PheThrTrpThrGlyAlaSerLysTyrAsnLeuAnGlyArgGluSerIleIleAsnPro 520
Db 2981 TACTCGTGAGCTGGAGCTACCAAGTACCACTTCATGGCAGAGCTCTCTGGTGAATCCG 3040
QY 521 GlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyVal 540
Db 3041 GGCCCGCCATCGCAAGCCACAGACGATGAAGAAAGTTTTTCTCAGAGCGGGTT 3100
QY 541 MetIlePheGlyGlySerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIle 560
Db 3101 CTCATCTTTGGGAAGCAAGCTCAGAGAAACAATGTGGACATGTGAAGAGTCTATGATT 3160
QY 561 ThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrVal 580
Db 3161 ACAGACGAGAGAGAAATACGAGCAACCAATCCGCTGGCTACGAGCAGTATGGTCTGTA 3220
QY 581 AlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGly 600
Db 3221 TCTACCACTCCAGAGAGGCAACAGACAGCAGCTACCGCAGATGTCAACACACAGGC 3280
QY 601 AlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAla 620
Db 3281 GTTCTTCCAGGATGCTTGGCAGGACAGAGATGTGTACTTTCAGGGGCCATCTGGCA 3340
QY 621 LysIleProHisThrAspGlyHisPheHisProSerProLeuMetClyGlyPheGlyLeu 640
Db 3341 AAGATTCACACACGACGACGACATTTTACCCCTCTCCCTCATGGGTGGATTCGACAT 3400
QY 641 LysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAla 660
Db 3401 AAACACCTCTCCACAGATCTCATCAAGACACACCCCGGTACCTCGGAATCTTCGACC 3460
QY 661 GluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSer 680
Db 3461 ACCTTCAGTCCGCAAGTTTGTCTCTTCATCACACAGTACTCTCCACGGGACAGGTTCAGC 3520
QY 681 ValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGln 700
Db 3521 GTGAGATCAGTGGGAGCTGCAGAGAGGAAACAGCAACGCTGGAAATCCGAAATTCAG 3580
QY 701 TyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAnGlyLeu 720
Db 3581 TACACTTCCAACTACAAAGTCTGTAAATGTGGACTTTTACTGTGGACACTAATGCGGTG 3640
```

```
QY 721 TyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 3641 TATTGAGAGCTCCGCCCAATTGGCAGCAGATACCTGACTCGTAATCTG 3688
RESULT 7
AAI66974
ID AAI66974 standard; DNA; 4679 BP.
XX AC AAI66974;
XX 11-FEB-2002 (first entry)
XX Adeno-associated virus 2 complete genomic sequence.
XX Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytosstatic;
XX inverted terminal repeat; nontropic; neuroprotective; antianemic; ITR;
XX antidiabetic; antitumour; gene therapy; adeno-associated virus; AAV; ds.
XX Adeno-associated virus 2.
XX Key Location/Qualifiers
FH CDS 321..2252
FT /*tag= a
FT /product= "Rep 68 protein"
FT /note= "contains introns; for coding sequence
FT join (321..1906, 2228..2252)"
FT CDS 321..2186
FT /*tag= b
FT /product= "Rep 78 protein"
FT 993..2252
FT /*tag= c
FT /product= "Rep 40 protein"
FT /note= "contains introns; for coding sequence
FT join (993..1906, 2228..2252)"
FT CDS 993..2186
FT /*tag= d
FT /product= "Rep 52 protein"
FT 2203..4410
FT /*tag= e
FT /product= "major coat protein VP1"
FT 2614..4410
FT /*tag= f
FT /transl_except= "(pos:2614..2616, aa: Met)"
FT /product= "major coat protein VP2"
FT 2809..4410
FT /*tag= g
FT /product= "major coat protein VP3"
XX WO200168888-A2.
XX 20-SEP-2001.
XX 13-MAR-2001; 2001WO-US07927.
XX 14-MAR-2000; 2000US-189110P.
XX (NEUR-) NEUROLOGIX INC.
XX Xiao W, During MJ;
XX WPI: 2001-596912/67.
XX P-PSDB; AAG65788, AAG65789, AAG65790, AAG65791, AAG65792, AAG65793,
XX AAG65794.
XX Recombinant viral vector useful in improving gene therapy in a subject,
XX and for increasing efficiency of entry into a cell, comprises a
XX chimeric capsid having one non-native amino acid sequence and a desired
XX transgene -
XX Disclosure; Page 46-47; 53pp; English.
XX The invention provides a recombinant viral vector (RVV) comprising a
XX chimeric capsid (I) having at least one non-native amino acid sequence,
CC
```

CC derived from a capsid protein domain of parvovirus (II), a virus (III),
 CC or their combination, and a transgene flanked 5' and 3' by inverted
 CC terminal repeat (ITR) sequences, derived from (II), (III), or their
 CC combination. The RRV is useful for improving gene therapy in a subject
 CC with a disorder, and for increasing the efficiency of entry into a cell,
 CC which involves producing (I) encapsulating a viral vector, and contacting
 CC a cell with RRV having (I) such that (I) binds to an attachment site on
 CC the cell surface and permits the vector to enter the cell efficiently.
 CC A pharmaceutical composition comprising RRV with (I) containing a
 CC transgene sequence associated with a disease or a disorder such that
 CC expression of the transgene would result in amelioration of the disease
 CC or disorder such as inherited neurological and metabolic diseases e.g.
 CC lysosomal storage diseases, Lesch-Nyhan syndrome, amyloid polyneuropathy,
 CC Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood
 CC e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic
 CC fibrosis, diabetes, diseases associated with hormone deficiencies,
 CC retinoblastoma and various types of neoplastic cells which include
 CC tumours especially central nervous system tumours, neoplasms, carcinomas,
 CC sarcomas, leukemias and lymphoma. The present sequence represents the
 CC complete genomic sequence of adeno-associated virus 2 which is used in
 CC the construction of a chimeric vector.

SQ Sequence 4679 BP; 1198 A; 1262 C; 1255 G; 964 T; 0 other;

Alignment Scores:

Pred. No.: 7,1e-240 Length: 4679
 Score: 3402.50 Matches: 613
 Percent Similarity: 90.22% Conservative: 51
 Best Local Similarity: 83.29% Mismatches: 71
 Query Match: 85.30% Indels: 1
 DB: 22 Gaps: 1

US-09-807-802A-13 (1-736) x AAI66974 (1-4679)

QY 1 MetAlaAlaAspGlyTyrLeuProAspTrpLeuGluAspAsnLeuSerGluGlyIleArg 20
 DB 2203 ATGGCTGCGCGATGGTTATCTTCAGATTGGCTCGAGGACACTCTCTCTGAAGGATATAAGA 2262
 QY 21 GluTrpTrpAspLeuLysProGlyAlaProLysProLysAlaAsnGlnLysGlnAsp 40
 DB 2263 CAGTGTGTGAAGCTCAAACTGGCCACCACCACCACCCGAGCGGCGGATAGGAC 2322
 QY 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
 DB 2323 GACAGCAGGGGCTTGTGCTTCTGGGTACAAGTACTCGACCTCGACCTTCAACCGACTCGAC 2382
 QY 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
 DB 2383 AAGGGAGACCGGTCAACGAGCGACGACCGCGGCGCTCGAGCAGCAAAAGCCTACGAC 2442
 QY 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
 DB 2443 CGCGAGCTCGACGCGGAGACCAACCGTACTTCAAGTACCAACGCGCGCGGAGTTT 2502
 QY 101 GlnGluArgGlnGlnAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln 120
 DB 2503 CAGGAGCGGCTTAAAGAAGATACGCTCTTTGGGGCAACCTCGGACGAGCAGTCTTCCAG 2562
 QY 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaPro 140
 DB 2563 GCGAAAAGAGGGTCTTGAACCTCTGGGCGCTGGTTGAGGACCTGTTAAGACCGGCTCG 2622
 QY 141 GlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerGlyIleGly 160
 DB 2623 GAAAAAAGAGCGCGGTAGGCACTCTCTGTGGAGCGAGACTCTCTCTCGGGAACCGGA 2682
 QY 161 LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
 DB 2683 AAGCGCGGCGCAGCAGCGCTGCAAGAAAAAGATTGAATTTGGTCAGACTGGAGACGACAG 2742
 QY 181 SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro 200
 DB 2743 TCAGTACCTTGACCCCGAGGCTCTCGGACAGCACCAGCAGCCCTCTCTGCTGCGAACT 2802

QY 201 ThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAspAsnGluGlyAlaAspGly 220
 DB 2803 AATACAGTGGCTACAGCAGTGGCGCACCAATGCGACACATAACAGAGGGCGCGACGGA 2862
 QY 221 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValile 240
 DB 2863 GTGGGTAAATCTCTCGGAAAATGGCAATTCGATTCACATGCGATGGGCGACAGAGTCAATC 2922
 QY 241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnHisLeuTyrLysGlnIle 260
 DB 2923 ACCACACGAGCCGAACTGGGCGCTCCACCTACCAACACCACTCTCAAAACAAAT 2982
 QY 261 SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrp 280
 DB 2983 TCCAGCCCAATCA---GGAGCCTCGAACGACAACTACTCTTTGGCTACAGCACCCCTTGG 3039
 QY 281 GlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu 300
 DB 3040 GGGTATTTTGACTTCAACAGATTCCACTGCGCACTTTTCCACCAGCTGACTGGCAAGACTC 3099
 QY 301 IleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGln 320
 DB 3100 ATCAACAACAACCTGGGATTCGACCCCAAGAGACTCAACTTCAAGCTCTTTAAACATTCA 3159
 QY 321 VallysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThr 340
 DB 3160 GTCAAGAGGTCAACGACGATGAGGTACGACGAGGATGGCCATAACCTTACCAGCAGC 3219
 QY 341 ValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGln 360
 DB 3220 GTTCAGGTGTTTACTGACTCGGAGTACGAGCTCCGCTACGCTCGGCTCGGCGCATCA 3279
 QY 361 GlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr 380
 DB 3280 GGATGCTCTCCCGCGTTCCTCCAGCAGAGCTTTCATGTTGGCCACAGTATGGATACCTCACC 3339
 QY 381 LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 400
 DB 3340 CTGAACAACGGAGTCAGGACAGTAGGACGCTCTTCACTTTACTGCTGGAGTACTTCTCT 3399
 QY 401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 420
 DB 3400 TCTCAGATGCTGCGTACCGGAAACACTTTTACCTTACGTACACTCTTTTGGAGGAGCTTCT 3459
 QY 421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
 DB 3460 TTCCAGACGACTAGCTCACAGCCAGAGTCTGGACCGCTCTCAGTGAATCTCTCATCGAC 3519
 QY 441 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAsp 460
 DB 3520 CAGTACCTGTATTACTTTGAGCAGAAACAACACTCCAAGTGGAAACCAACCCAGCAGTCAAGG 3579
 QY 461 LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuPro 480
 DB 3580 CTTCAGTTTTCTCAGCGCGGAGCGAGTGACATCTCGGACCCAGTCTAGGACCTGGCTTCT 3639
 QY 481 GlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsn 500
 DB 3640 GGACCTGTATACCCGACGAGCGAGTATCAAGACATCTCGGAGATACCAACACACAGTGA 3699
 QY 501 PheThrTrpThrGlyAlaSerLysThrAsnLeuAsnGlyArgGluSerIleIleAsnPro 520
 DB 3700 TACTCGTGAGCTGGAGCTTACCAGTACCCTCAATGGCAGAGACTCTCTGCTGAATCCG 3759
 QY 521 GlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyVal 540
 DB 3760 GGGCGGCGCATGGCAAGCCACAGGACGATGAAGAAAGTTTTTCTCTCAGACGCGGGT 3819
 QY 541 MetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIle 560
 DB 3820 CTCATCTTTGGAGACGAGGCTCAGAGAAACAAATGTGGACATTTGAAAGGTCTATGATT 3879

| | | | |
|----|------|---|------|
| Qy | 561 | ThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrVal | 580 |
| Db | 3880 | ACAGACGAAGAAATCAGACAACCAATCCCGTGGCTACGGAGCAGTAGTATGGTTCTGTA | 3939 |
| Qy | 581 | AlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGly | 600 |
| Db | 3940 | TCTACCAACCTCCAGAGAGCGACAGACAGCAGCTACCGCAGATGTCAACACACACAGGC | 3999 |
| Qy | 601 | AlaLeuProGlyMetValTrpGlnAspArgAspValTyrIleuGlnGlyProIleTrpAla | 620 |
| Db | 4000 | GTTCTTCCAGGCATGGTCTGGCAGGACAGAGATGTGTACCTTCAGGGGGCCCATCTGGGCA | 4059 |
| Qy | 621 | LysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeu | 640 |
| Db | 4060 | AGANTTCCACACAGCGAGCGACATTTTCAACCCCTCTCCCTCATGGGTGGATTCGGACTT | 4119 |
| Qy | 641 | LysAsnProProGlnIleLeuIleGlyAsnThrProValProAlaAsnProProAla | 660 |
| Db | 4120 | AAACACCTCTCTCCACAGATTTCTCATCAAGAACAACCCCGGTACCTGCGAAATCTCTCGACC | 4179 |
| Qy | 661 | GluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSer | 680 |
| Db | 4180 | ACCTTCAGTGGCGCAAGTTTGGTTCTTCTCATCACAGTACTCCACGGAGCAGGTGACG | 4239 |
| Qy | 681 | ValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGln | 700 |
| Db | 4240 | GTGGAGATCGAGTGGGAGCTGCAGAGGAAAAACAGCAAAACGCTGGAATCCCGAAATTCAG | 4299 |
| Qy | 701 | TyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeu | 720 |
| Db | 4300 | TACACTCCAACTACAACAGTCTGTAAATGTGGACTTTACTTGGGACACTAATGGCGTG | 4359 |
| Qy | 721 | TyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu | 736 |
| Db | 4360 | TATTCAAGCTCCGCCATTGGCACGACATCTGACTGCTGTAATCTG | 4407 |

RESULT 8

| | |
|---------|--|
| AA23750 | |
| ID | AAF23750 standard; DNA; 4679 BP. |
| XX | |
| AC | AAF23750; |
| XX | |
| DT | 28-MAR-2001 (first entry) |
| XX | |
| DE | AAV2 DNA sequence. |
| XX | |
| KW | AAV2; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS; |
| KW | atherosclerosis; sickle cell anaemia; thalassaemia; |
| KW | blood clotting disorder; diabetes; ss. |
| XX | |
| OS | Adeno associated virus. |
| XX | |
| FN | US6156303-A. |
| XX | |
| PD | 05-DEC-2000. |
| XX | |
| PF | 11-JUN-1997; 97US-0873168. |
| XX | |
| PR | 11-JUN-1997; 97US-0873168. |
| XX | |
| PA | (UNIW) UNIV WASHINGTON. |
| XX | |
| PI | Russell DW, Rutledge EA; |
| XX | |
| DR | WPI; 2001-060164/07. |
| XX | |
| PT | Adeno-associated virus serotype 6 and viral vector derived from it for |
| PT | gene therapy of cystic fibrosis, cancer, acquired immunodeficiency |
| PT | syndrome, sickle cell anemia, thalassaemia and diabetes - |
| XX | |
| PS | Claim 7; Fig 1; 50pp; English. |
| XX | |
| CC | The present invention relates to adeno-associated virus serotypes. The |

261 SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrp 280
2983 TCAGCCCAATCA---GGAGCCTCGAACGACATCACTACTTTGGCTACAGCACCCCTTGG 3039
281 GlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu 300
3040 GGGATATTTTGGCTTCAACAGATTCCACTGCCACTTTTACCACGCTGACGGCAAGACTC 3099
301 IleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheHisValPheAsnIleGln 320
3100 ATCAACACCACTGGGGATTCGAGCCCAAGAGACTCAACTTCAAGCTCTTTAAACATTCAA 3159
321 ValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThr 340
3160 GTCAAGAGGTCAAGCAGATGACGTACGACGAGATTGCCATTAACCTTACACGACG 3219
341 ValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGln 360
3220 GTTCAGGTGTTTACTGACTCGAGTACCACTCCGCTCGCTCGGCTCGGCGCATCAA 3279
361 GlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr 380
3280 GGATGCTCCCGCGTTCCAGCAGACGCTTTCATGGTGCACAGTATGGATACCTCACC 3339
381 LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 400
3340 CTGAACACCGGAGTACGACGATGAGCAGCTCTTCATTTTACTGCTCGGAGTACTTCT 3399
401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 420
3400 TCTCAGATGCTGGTACCGGAAACAACCTTTACCTTCAGCTACACTTTTTCAGGACGTTCT 3459
421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProIleLeuAsp 440
3460 TTCACAGCAGCTACGCTACAGCCAGAGTCTGACCGCTCTCATGAATCTCTCATCGAC 3519
441 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAsp 460
3520 CAGTACTGTATTTACTTGGACAGACAAACACACTCCCAAGTGGACACACCGCAGTCAAG 3579
461 LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuPro 480
3580 CTTTCAGTTTCTCAGCGCGGAGCGAGTGACATTTCCGGACCGAGTCTAGGAACTGGCTTC 3639
481 GlyProCysTyrArgGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsn 500
3640 GGACCTGTGTACCGCAGCAGCAGTCAACAGACATCTCCGATTAACACACACAGTGAA 3699
501 PheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleLeuAsnPro 520
3700 TACTCGTGGACTGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTGGTGAATCCG 3759
521 GlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyVal 540
3760 GGCCCGGCCCATGGCAAGCCCAAGGACGATGAAGAAAAGTTTTTCTCAGACGGGGTT 3819
541 MetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIle 560
3820 CTATCTTTGGAGAGAGGCTCAGAGAAACAATGTGGACATTTGAAGGTCATGATT 3879
561 ThrAspGluGluLysLysAlaThrAsnProValAlaThrGluArgPheGlyThrVal 580
3880 ACAGACGAGAGGAAATCAGACAAACCAATCCGCTGGCTACGAGCAGTATGTTCTGTGA 3939
581 AlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGly 600
3940 TCTACCAACTCTCCAGAGAGGCAACACAGACAGCTACCGCAGATGTCAACACACCAAGGC 3999
601 AlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAla 620
4000 GTTCTTCAGGCATGTCTGGCAGCAGAGATGTGTACCTTCAGGGGCCCATCTGGCA 4059
621 LysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeu 640

4060 AAGATTCCACACACGAGCGACATTTTCCACCCCTCTCCCTCATGGGTGGATTGGACTT 4119
641 LysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAla 660
4120 AAACACCCCTCTCCACAGATTCTCATCAAGAACACCCCGGTACCTGCGAATCTTCGACC 4179
661 GluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSer 680
4180 ACCTTCAGTCGGCAAGTTTCTTCTTATCACAGTACTCCACGGGACAGGTGACG 4239
681 ValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTyrAsnProGluValGln 700
4240 GTGGAGATCGAGTGGAGCTCGAAGGAAACAGCAACGCTGGAATCCGGAATTCAG 4299
701 TyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnGlyLeu 720
4300 TACACTTCCAACTACAACAGTCTGTTAATGTGACTTTTACTGTGGACACTAATGGCGTG 4359
721 TyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
4360 TATTGAGAGCTCGCCCATTTGGCACCAGATACCTGACTCTGTAATCTG 4407
RESULT 9
ABK89694
ID ABK89694 standard; DNA; 4679 BP.
XX AC ABK89694;
XX 05-NOV-2002 (first entry)
XX Adeno-associated virus 2 (AAV2) vector.
XX Adeno-associated virus 2 vector; AAV2; ds; cyclic; circular; cancer;
KW VPI capsid; heparin-sulphate proteoglycan; vaccine; immune response;
XX ovarian cancer.
XX Adeno-associated virus 2.
XX Key Location/Qualifiers
FH 2203..4410
FT /*tag= a
FT /product= "Adeno-associated virus 2, VPI capsid
FT protein"
FT CDS 2614..4410
FT /*tag= b
FT /product= "Adeno-associated virus 2, VP2 capsid
FT protein"
FT /transl_except= (pos:2614..2616, aa:Met)
FT 2809..4410
FT /*tag= c
FT /product= "Adeno-associated virus 2, VP3 capsid
FT protein"
XX WO200253703-A2.
XX 11-JUL-2002.
XX 04-JAN-2002; 2002WO-US00152.
XX 05-JAN-2001; 2001US-260124P.
XX (CHIL-) CHILDRENS HOSPITAL INC.
XX Bartlett JS;
XX WPI; 2002-583608/62.
DR P-PSDB; AAU98974, AAU98975, AAU98976.
XX New adeno-associated virus vector comprises a biotinylated capsid or
PT capsid protein with an amino acid insertion in the VPI capsid, useful
PT as a vaccine or for transferring a therapeutic peptide to a cancer cell
PT

XX Disclosure; Page 46-48; 57pp; English.
 XX
 CC The invention relates to an adeno-associated virus (AAV) vector (I)
 CC comprising a biotinylated capsid or capsid protein (II) with an amino
 CC acid insertion following the capsid amino acid at position 139, 161, 588
 CC or 657 in the VP1 capsid. The AAV vector comprises a capsid protein
 CC containing one or more amino acid insertions that ablate the ability of
 CC the vector to bind heparin-sulphate proteoglycan and allow the vector to
 CC use a cellular receptor not used by wild type AAV. Modified (I) are
 CC useful as vaccines to elicit immune responses to amino acids, where the
 CC response can be protective and/or therapeutic. (I) may be used to
 CC transfer a therapeutic peptide to a cancer cell, particularly to an
 CC ovarian cancer cell. The present sequence represents the adeno-associated
 CC virus 2 (AAV2) vector used in to make modified AAV2 vectors.
 XX
 SQ Sequence 4679 BP; 1198 A; 1262 C; 1255 G; 964 T; 0 other;

Alignment Scores:

Pred. No.: 7.1e-240 Length: 4679
 Score: 3402.50 Matches: 613
 Percent Similarity: 90.22% Conservative: 51
 Best Local Similarity: 83.29% Mismatches: 71
 Query Match: 85.30% Indels: 1
 DB: 24 Gaps: 1

US-09-807-802A-13 (1-736) x ABK99694 (1-4679)

Qy 1 MetAlaAlaAspGlyTyrLeuProAspTrpLeuGluAspAsnLeuSerGluGlyLeuArg 20
 Db 2203 ATGGCTCGCATGGTATCTTCCAGATTGGCTCGAGGACATCTCTCTGAAGGAATAGA 2262
 Qy 21 GluTrpTrpAspLeuLeuProGlyAlaProLysProLysAlaAsnGlnGlnLysGlnAsp 40
 Db 2263 CAGTGGTGAAGCTCAAACTGGCCACCACCACCACCAAGCCGCGAGCGCATGAAGGAC 2322
 Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
 Db 2323 GACAGCGGGGCTTGTCTCTCGGTGACAGTACCTCGGACCTTCAACGGACTCGAC 2382
 Qy 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
 Db 2383 AAGGGAGAGCGGTCAACGAGGAGAGCGCGCGGCTCTCGAGCAGCAAGGCTTACGAC 2442
 Qy 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
 Db 2443 CGGCAGCTCGACAGCGGAGACAACCGGTACCTCAAGTACAACCAACCGCGCGGAGTTT 2502
 Qy 101 GlnGluArgLeuGlnGlnAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln 120
 Db 2503 CAGGAGCGGCTTAAAGAGATACGCTTTTGGGGCAACCTCGGACGAGCAGTCTTCCAG 2562
 Qy 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGlyAlaLysThrAlaPro 140
 Db 2563 GCGAAAAGAGGGTCTTGAACCTCTGGGCTGGTTGAGGAACCTGTTAAGCGGCTCCG 2622
 Qy 141 GlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSerGlyLeuGly 160
 Db 2623 GGAATAAAGAGCGCGGTAGACACTCTCTGTGGAGCCAGACTCTCTCTCGGGAACCGGA 2682
 Qy 161 LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
 Db 2683 AAGCGGCGCCAGCAGCCTTGAAGAAAGATGAATTTTGTTCAGACTGGAGAGCGCAGAC 2742
 Qy 181 SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro 200
 Db 2743 TCAGTACTGTACCCCGACGCTCTCGGACAGCCAGCAGCGCCCTCTGTGTCTGGGAAC 2802
 Qy 201 ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnGlnGlyAlaAspGly 220
 Db 2803 AATACGATGGCTTACAGGCAGTGGCCGCAATGGCAGACAATAACAGGCGCGCGCAGGA 2862
 Qy 221 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIle 240

Db 2863 GTGGTAATTCCTCGGGAATTTGGCATTTCCACATGGATGGGCGCAGAGTCATC 2922
 Qy 241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIle 260
 Db 2923 ACCACAGAGCCCGAACCTGGGCGCTCCACCTACCAACCAACCTCTACAAACAAT 2982
 Qy 261 SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrp 280
 Db 2983 TCCAGCCCAATCA--GGAGCGCTCGAACGACAATCACTACTTTGGCTACAGCACCCCTGG 3039
 Qy 281 GlyTyrPheSerPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu 300
 Db 3040 GGGTATTTTACATTCAACAGATTCCACTGGCCACTTTTCCACCGCTGACTGGCAAGACTC 3099
 Qy 301 IleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGln 320
 Db 3100 ATCAACAACAACCTGGGATTTCCGACCCCAAGAGACTCAACTTCAAGCTCTTTAACAATCAA 3159
 Qy 321 ValLysGluValThrThrAsnAspGlyValThrThrIleAlaAlaAsnLeuThrSerThr 340
 Db 3160 GTCAAGAGGTCACGCGAGAATGACGATACGACGACGATTCGCAATAACCTTACCAGACG 3219
 Qy 341 ValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGln 360
 Db 3220 GTTCAGGTGTTTACTGACTCGGAGTACAGCTCCCGTACGCTCTCGGCTCGGCCATCAA 3279
 Qy 361 GlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr 380
 Db 3280 GGATGCTCTCCGCGCTTCCAGCAGACGCTCTCATGTGTGCCACAGTATGATACCTCAC 3339
 Qy 381 LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 400
 Db 3340 CTGAACAACGGGATGACGAGTAGGACGCTCTTCAATTTACTGCTCGGTAGTACTTCTCT 3399
 Qy 401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 420
 Db 3400 TCTCAGATGCTGGTACCGGAAACAACCTTTACCTTTCAGCTACACTTTTGGAGGAGCTTCT 3459
 Qy 421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
 Db 3460 TTCCACAGCAGCTACGCTCACGCCAGAGCTGAGCGCTCTCATGAATCTCTCATCGAC 3519
 Qy 441 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAsp 460
 Db 3520 CAGTACCTGTATTACTTTGAGCAGACAACAACCTCAAGTGGAAACCAACCGCAGCTCAAG 3579
 Qy 461 LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuPro 480
 Db 3580 CTTCAGTTTTCTCAGGCGGAGGAGTGACATTCGGGACCGACTCTAGAACTGGCTTCTCT 3639
 Qy 481 GlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsn 500
 Db 3640 GGACCTCTGTACCGCCAGCAGCAGTATCAAGACATCTCGGATACAACAACAGTGAA 3699
 Qy 501 PheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnPro 520
 Db 3700 TACTCGTGGACTGGAGTACCAAGTACCACTCAATGGCAGAGACTCTCTGGTGAATCCG 3759
 Qy 521 GlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyVal 540
 Db 3760 GGCCCGGCCATGGCAAGCCCAAGGAGCATGAAGAAAGTTTCTCTCAGAGCGGGGTT 3819
 Qy 541 MetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIle 560
 Db 3820 CTCACTTTGGGAAGCAAGGCTCAGAGAAAACAATAATGTGGACATTTGAAAGGTTCATGAT 3879
 Qy 561 ThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrVal 580
 Db 3880 ACAGACCAAGAGGAAATCAGGACAACAATCCCGTGGCTACGGAGCAGTATGGTCTCTGTA 3939
 Qy 581 AlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGly 600

Db 3940 TCTACCACTCCAGAGGCAACACAGACAGCAGCTACCGCAGATGTCAACACACAGGC 3999
 QY 601 AlaLeuProGlyMetValTrpGlnAspArgAspValTrpGlnGlyProIleTrpAla 620
 Db 4000 GTTCTTCCAGGCATGCTCTGGCAGGACAGAGATGTGTACTTTCAGGGGCCCATCTGGGCA 4059
 QY 621 LysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeu 640
 Db 4060 AAGATTCCACACAGGACGGACATTTTCACCCCTCTCCCTCATGGGTGGATTCGGACTT 4119
 QY 641 LysAsnProProGlnIleLeuLysAsnThrProValProAlaAsnProAla 660
 Db 4120 AAACACCTCTCTCCAGATTCTCATCAAGACACCCCGTACTTGGATCTTCGACC 4179
 QY 661 GluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSer 680
 Db 4180 ACCCTTCAGTGCAGCAAGTTTGTCTCTTCATCACACAGTACTCCACGGGACAGGTCTAGC 4239
 QY 681 ValGluIleGluTrpGluLeuGlnLysGluAsnSerLysAsnThrProGluValGln 700
 Db 4240 GTGGAGATCGAGTGGGAGCTGCAGAGGAAACAGCAACGCTGGATCCCGAATTCAG 4299
 QY 701 TyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeu 720
 Db 4300 TACACTTCCAACTACAAAGTCTGTATATGTGGACTTTTACTGTGGACACTAATGGCGTG 4359
 QY 721 TyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
 Db 4360 TATTCAGAGCCTGGCCCCATTGGCACCATATCTGACTCGTAATCTG 4407

RESULT 10

ABV76133
 ID ABV76133 standard; DNA; 4679 BP.
 AC ABV76133;
 XX
 DT 07-MAR-2003 (first entry)
 XX
 DE Adeno associated virus type 2 coat protein nucleic acid.
 XX
 KW Gene therapy; vector; hepatitis B virus; cardiovascular disease;
 KW heart; cardiac; vasotropic; antiarrhythmic; antiarteriosclerotic;
 KW coat protein; gene; ss.
 XX
 OS Adeno associated virus type 2.
 XX
 PN W0200287594-A1.
 XX
 PD 07-NOV-2002.
 XX
 PF 30-APR-2002; 2002WO-US13644.
 XX
 PR 30-APR-2001; 2001US-287423P.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Chien KR, Hoshijima M;
 XX
 DR WPI; 2003-111844/10.
 XX
 PT Novel non-viral vector comprises vesicular membrane with hepatitis B
 PT envelope protein with cardiac targeting sequence, and nucleotide
 PT sequence for gene therapy useful for treating, e.g., heart failure,
 PT arrhythmia and atherosclerosis -
 XX
 PS Disclosure; Page 24-27; 53pp; English.
 XX
 CC The present sequence is that of an adeno-associated virus (AAV) type
 CC 2 coat protein nucleic acid. The invention provides a non-viral
 CC vesicle vector for the delivery of nucleic acid to various cardiac
 CC cell types. The vesicle vector contains the hepatitis B virus
 CC envelope protein in which at least part of the liver targeting
 CC sequence is deleted and replaced with a specific cardiac cell

CC targeting sequence. For example, the loop IV region of the AAV
 CC binds to heparin sulfate proteoglycans on the surface of
 CC cardiomyocytes, and can be used to target the vector to
 CC cardiomyocytes. The vesicle vector can be delivered intravenously
 CC or intra-arterially rather than by more invasive methods such as
 CC direct cardiac injection. It can be used to deliver gene products
 CC to replace or enhance expression of proteins for treatment of heart
 CC failure, arrhythmia, reperfusion injury, atherosclerosis, to
 CC promote angiogenesis, etc. The vesicles are highly stable and can
 CC be produced in large quantities, making them ideal for gene therapy.

XX Sequence 4679 BP; 1198 A; 1262 C; 1255 G; 964 T; 0 other;

Alignment Scores:

Pred. No.: 7,18-240 Length: 4679
 Score: 3402.50 Matches: 613
 Percent Similarity: 90.22% Conservativity: 51
 Best Local Similarity: 83.29% Mismatches: 71
 Query Match: 85.30% Indels: 1
 DB: Gaps: 1

US-09-807-802A-13 (1-736) x ABV76133 (1-4679)

QY 1 MetAlaAlaAspGlyTyrLeuProAspTrpLeuGluAspAsnLeuSerGluGlyIleArg 20
 Db 2203 ATGCTGCGGATGGTTATCTTCAGATTGGCTCGAGGACACTCTCTGNAAGGAATAAGA 2262
 QY 21 GluTrpTrpAspLeuLysProGlyAlaProLysProLysAlaAsnGlnGlnGlnAsp 40
 Db 2263 CAGTGTGGAAAGCTCAAACTGGGCCACCACCAACCAAGCCCGCAGCATGAAGAC 2322
 QY 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTrpLeuGlyProPheAsnGlyLeuAsp 60
 Db 2323 GACAGCAGGGGTCTTGTCTCTCTGGGTACAGTACTCTCGGACCTTCAACGGACTCGAC 2382
 QY 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
 Db 2383 AAGCGAGAGCGGTCAACGAGGACAGCGCCGCGCTCGAGCAGCAACAGCTTCGAC 2442
 QY 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
 Db 2443 CGCGAGCTCGACGCGGAGACACCCGTACCTCAAGTACAACCCAGCCGCGCGGAGTTT 2502
 QY 101 GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln 120
 Db 2503 CAGGAGCCCTTAAAGAAAGATACGTCTTTTGGGGCAACCTTCGACGAGCAGTCTTCAG 2562
 QY 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaPro 140
 Db 2563 GCGAAAAAGAGGGTTCTTGAACCTCTGGGCTGGTGTAGGAACCTGTTAAGACGCGCTCG 2622
 QY 141 GlyLysLysArgProValGluLysSerProGlnGluProAspSerSerSerGlyIleGly 160
 Db 2623 GGAATAAAGAGCGGTAGAGCACTCTCTGTGGAGCCAGACTCTCTCTCGGAAACCGGA 2682
 QY 161 LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
 Db 2683 AAGCGGCGCCAGCAGCTGCAAGAAAAGATTGAATTTTGTTCAGACTGGAGACGACAG 2742
 QY 181 SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro 200
 Db 2743 TCAGTACCTGACCCCGCAGCTCTCGACAGCCAGCCAGCAGCCCTCTGTCTCGGAACT 2802
 QY 201 ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGly 220
 Db 2803 AATACGATGGCTACAGGCGAGTGGCGCCCAATGGCAGACAATACAGGGGCCCGACGGA 2862
 QY 221 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIle 240
 Db 2863 GTGGGTAAATCTCTCGGAAATTTGGCATTCGATTCACATGATGGGCGACAGAGTCATC 2922
 QY 241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIle 260

Db 2923 ACCACGAGCCCGAAGCTGGGCGCTGCCACCTACCAACACCACTCTCAACAACAAATT 2982
Qy 261 SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTyr 280
Db 2983 TCCAGCCCAATCA---GGAGCTCTGACGACGACATCACTACTTTGGCTACAGCACCCCTTGG 3039
Qy 281 GlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTyrGlnArgLeu 300
Db 3040 GGGTATTTGACTTCAACAGATTCCACTGCGCACTTTTCCACCAGCTGACTGCGCAAGACTC 3099
Qy 301 IleAsnAsnAsnTyrGlyPheArgProLysArgLeuAsnPheTyrLeuPheAsnIleGln 320
Db 3100 ATCAACAACTCGGGATTCGAGCCCAAGAGACTCAACTTCAAGCTCTTTAAACATTCAA 3159
Qy 321 ValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThr 340
Db 3160 GTCAAGAGGTCCACGCAATGACGGTACGAGCAGGATTCGCAATACCTTACCAGCAGC 3219
Qy 341 ValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGln 360
Db 3220 GTTCAGGTGTTTACTGACTCGGAGTACCAAGCTCCGCTACGTCTCGGCTCGGCGCATCAA 3279
Qy 361 GlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr 380
Db 3280 GGATGCTTCCCGCGTTCCTCCAGCAGAGCTTTCATGTTGGTCCACAGTATGGATACCTCAC 3339
Qy 381 LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 400
Db 3340 CTGAACAACGGAGTCAGGCACTAGGAGCTCTTCACTTTTACTGCTGGAGTACTTTCCT 3399
Qy 401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 420
Db 3400 TCTCAGATGCTGCGTACCGGAAACAACTTTTACCTTCAGTACACTTTTGGAGGAGTTCCT 3459
Qy 421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
Db 3460 TTCACAGACCTAGCTTCACAGCAGAGTCTGGACCGCTCATGAATCTCTCATCGAC 3519
Qy 441 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAsp 460
Db 3520 CAGTACCTGTATTACTTGGCAGAGAAACAACTCCAGTGGGAACCAACCCAGCAGTCAAGG 3579
Qy 461 LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTyrLeuPro 480
Db 3580 CTTCAAGTTTCTCAGCGCGGAGCGAGTGCATCTCGGACCACTAGGAACTGGCTTCCT 3639
Qy 481 GlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsn 500
Db 3640 GGACCTGTATCCCGCCAGCAGGAGTATCAAGACATCTCGCGATACACACACAGTGAA 3699
Qy 501 PheThrTyrThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleAsnPro 520
Db 3700 TACTCGTGGAGTGGAGTACCAAGTACCACTCAATGGCAGAGCTCTCTGGTGAATCCG 3759
Qy 521 GlyThrAlaMetAlaSerHisLysAspGluAspLysPhePheProMetSerGlyVal 540
Db 3760 GGCCCGGCGCATGGCAAGCACAAGACGATGAAGAAAGTTTTCTCTCAGAGCGGGT 3819
Qy 541 MetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIle 560
Db 3820 CTCATCTTTGGAGAGAGGCTCAGAGAAACAAATGTGGACATTTGAAGAGTCATGAT 3879
Qy 561 ThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrVal 580
Db 3880 ACAGACGAAGAGGAATCAGGACCAACCAATCCCGTGGCTACGGAGCAGTATGGTCTGTA 3939
Qy 581 AlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGly 600
Db 3940 TCTACCACTCCAGAGAGGCAACACAGACGAGCTACCGAGATGTCAACACACAGGCG 3999
Qy 601 AlaLeuProGlyMetValTyrGlnAspArgAspValTyrLeuGlnGlyProIleTyrAla 620
Db 4000 GTTCTTCCAGGATGGTCTGGCAGGACAGAGATGTGTACCTTACGGGGCCCATCTGGCA 4059

Qy 621 LysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeu 640
Db 4060 AAGATTCCACACACGACGACGACATTTTACCCCTCTCCCTCATGGGTGGATTCCGACTT 4119
Qy 641 LysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAla 660
Db 4120 AAACACCCCTCTCCACAGATTCTCATCAAGAACACCCCGGTACTCTGCGAATCTCTCGACC 4179
Qy 661 GluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSer 680
Db 4180 ACCTTCAGTCCGCAAGATTGCTTCTTCTCATCACAGTACTTCCACGGACAGGTCTCAGC 4239
Qy 681 ValGluIleGluTyrGluLeuGlnLysGluAsnSerLysArgTyrAsnProGluValGln 700
Db 4240 GTGGAGATCAGTGGGAGCTGCAGAGGAAACACAGAAACGCTGGAAATCCGAAATTCAG 4299
Qy 701 TyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeu 720
Db 4300 TACACTTCCAACACTACAACAGTCTGTTAATGTGGACTTTTACTGTGGACACTAATGGCGTG 4359
Qy 721 TyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 4360 TATTCAGAGCTCGCCCATTTGGACCAAGATACCTGACTGTAATCTG 4407

RESULT 11
AAH26326
ID AAH26326 standard; DNA; 7557 BP.
XX AC AAH26326;
XX DT 02-OCT-2001 (first entry)
XX DE Adeno-associated virus packaging plasmid pfloxAAV.
XX KW AAV; vector; pfloxAAV2; packaging cell line; gene therapy; ds.
XX OS Adeno associated virus.
XX FN W0200155361-A2.
XX PD 02-AUG-2001.
XX PF 26-JAN-2001; 2001WO-US02709.
XX PR 26-JAN-2000; 2000US-0178536.
XX PA (CHIR) CHIRON CORP.
XX PI Hardy SF;
XX WPI; 2001-483239/52.
XX PT Producing recombinant adeno-associated virus (rAAV) vector, by stably
XX FT infecting eukaryotic host cell with rAAV vectors, helper adenovirus or
XX PS herpes virus, recombinant adenovirus or herpes vectors -
XX Example 1; Page 54-57; 63pp; English.
XX The present sequence is that of adeno-associated virus (AAV)
XX packaging plasmid pfloxAAV2. The following AAV packaging plasmids
XX were prepared: pK5repcep, a non-replicating control plasmid
XX containing rep and cap genes but no inverted terminal repeats
XX (ITRs); pAV2 (see AAH26325), containing a wild-type AAV genome; and
XX pfloxAAV, a version of pAV2 with 2 loxp sites inserted such that
XX flank the rep and cap genes and thus separate the ITRs from these
XX genes. Each was combined with pCMV GFP (see AAH26324), a plasmid
XX with a green fluorescent protein expressing recombinant AAV (rAAV)
XX genome, and used to transfect 293 or Cre8 cells. After 6 hr, the
XX medium was changed and wild-type adenovirus type 5 was added.
XX Virus particles were harvested 3 days later. Plasmid pAV2
XX produced mostly AAV and a low yield of GFP vector with no effect
XX of Cre recombinase. pfloxAAV packaged as much GFP vector as the

CC control plasmid and while the amount of GFP vector was not
 CC affected by Cre recombinase, the amount of floxA^{AV} in the product
 CC was reduced to 1/10 of the 293 value by the action of Cre
 CC recombinase. In 293 cells it was evident that the GFP vector did
 CC not compete with wild-type AAV for replication or packaging, while
 CC the GFP vector was able to complete against floxA^{AV}. Methods
 CC and compositions are provided for producing rAAV vector particles
 CC by: (a) introducing into a host cell (i) p_{CMV}AAV, (ii) a
 CC recombinant viral vector encoding plasmid, and (iii) a plasmid
 CC encoding herpesvirus, cytomegalovirus or adenoviral functions, or
 CC a herpesvirus, cytomegalovirus or adenovirus itself, to produce
 CC floxA^{AV} particles and rAAV particles; and (b) introducing into a
 CC second host cell (i) the rAAV particles or (a), (ii) a vector
 CC that directs expression of Cre, and (iii) a vector which directs
 CC expression of herpes virus, cytomegalovirus or adenovirus helper
 CC functions, such that rAAV vector particles are produced. The
 CC vectors are useful for in vivo or in vitro gene therapy and also
 CC for in vitro recombinant protein production.

XX
 SQ Sequence 7557 BP; 1970 A; 1940 C; 1953 G; 1694 T; 0 other;

Alignment Scores:

Pred. No.: 1.31e-239 Length: 7557
 Score: 3402.50 Matches: 613
 Percent Similarity: 90.2% Conservative: 51
 Best Local Similarity: 83.2% Mismatches: 71
 Query Match: 85.3% Indels: 1
 DB: 22 Gaps: 1

US-09-807-802a-13 (1-736) x AAH26326 (1-7557)

```

QY 1 MetAlaAlaAspGlyTyrLeuProAspTrrPLeuGluAspAsnLeuSerGluGlyIleArg 20
DB 2231 ATGGCTGCGCGATGGTATCTTCAGATTGGCTCGAGGACACTCTCTCTGAAGGAATAAGA 2290

QY 21 GluTrrPrrAspLeuPrrGlyAlaPrrAspPrrAspPrrAspPrrAspPrrAspPrrAsp 40
DB 2291 CAGTGGTGAAGCTCAAACTTGGCCACCCACCACCCAGGCGGCGGCGGCGGCGGCGGCGG 2350

QY 41 AspGlyArgGlyLeuValLeuProGlyTyrLeuPrrAspPrrAspPrrAspPrrAspPrrAsp 60
DB 2351 GACAGCAGGCGGTCTTGTCTTCTGGGTACAAGTACTCTCGACCTTCAACGCGACTCGAC 2410

QY 61 LysGlyGluProValAsnAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 80
DB 2411 AAGGGAGACCGGTCAACAGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2470

QY 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHiAlaAlaAlaGluPhe 100
DB 2471 CGGCGAGCTCGACAGCGGAGACAACCGGTACTCTCAAGTACAACCCAGCGCGCGGAGTTT 2530

QY 101 GlnGluArgGlnGlnAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln 120
DB 2531 CAGGAGCGCTTAAAGAAAGATACGTCTTTTGGGGCAACCTCGGACGAGCACTCTTCCAG 2590

QY 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaPro 140
DB 2591 GCGAAAAAGAGGGTCTTGAACCTCTGGGCTGGTTGAGGAACTCTGTTAAGACGCTCG 2650

QY 141 GlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerGlyLysGly 160
DB 2651 GCGAAAAAGAGGGTCTTGAACCTCTGGGCTGGTTGAGGAACTCTGTTAAGACGCTCG 2710

QY 161 LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
DB 2711 AAGCGCGCGCGAGCGCTCGAAGAAAGATTAATTTTGGTCAGACTCGGACGCGAGAC 2770

QY 181 SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro 200
DB 2771 TCAGTACCTTGACCCCGAGCTCTCGGACGACGACGACGACGACGACGACGACGACGAC 2830

QY 201 ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGly 220

```

```

DB 2831 AATACGATGGCTACAGCGAGTGGCGCACCAATGCGACACAATACGAGGGCGCGCGGCA 2890
QY 221 ValGlyAsnAlaSerGlyAsnTrrPrrHisCysAspSerThrTrrPrrLeuGlyAspArgValle 240
DB 2891 GTGGGTAAATCTCGGAAATTTGGCATTTCCACATGATGGGCGACAGAGTCAATC 2950
QY 241 ThrThrSerThrArgThrTrrAlaLeuProThrTrrAsnAsnHisLeuTrrLysGlnIle 260
DB 2951 ACCACGAGCAGCCGAACTGGGCCCTGCCACCTTACCAACACCACTTACCAACAATTT 3010
QY 261 SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTrrPrrHisTrrSerThrProTrrP 280
DB 3011 TCCAGCCCAATCA---GGAGCTCGAACGACAATCACTACTTTTGGCTACAGCACCCCTGG 3067
QY 281 GlyTrrPrrAspPrrAsnAspPrrPheHisCysHisPheSerProArgAspTrrPrrArgLeu 300
DB 3068 GGGTATTTTGAATTTCAACAGATTCCACTGGCCACTTTTCCACCAGCTGATGGCAAGACTC 3127
QY 301 IleAsnAsnAsnTrrPrrGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGln 320
DB 3128 ATCAACAACAACCTGGGATTTCCGACCCCAAGAGACTCAACTTCAAGCTCTTTTAACTTCA 3187
QY 321 ValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThr 340
DB 3188 GTCAAGAGGTCACGCGAGATGACGGTACGACGACGATTTGCCAATAACCTTACCAGCAGC 3247
QY 341 ValGlnValPheSerAspSerGluTrrGlnLeuProTrrValLeuGlySerAlaHisGln 360
DB 3248 GTTCAGGTGTTTACTGACTCGGAGTACGAGCTCCGCTCCGCTCCGCGCTCGGCGATCAA 3307
QY 361 GlyCysLeuProProPheProAlaAspValPheMetIleProGlnTrrGlyTrrLeuThr 380
DB 3308 GGATGCTCTCCCGCGTTCCTCCAGCAGACGCTTTCATGTGGCCACAGTAGTAGGATACTCAC 3367
QY 381 LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTrrCysLeuGluTrrPhePro 400
DB 3368 CTGAACAACCGGAGTCAAGGAGTAGGAGCGCTCTTCAATTTACTGCTGGAGTACTTTCCT 3427
QY 401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTrrThrPheGluGluValPro 420
DB 3428 TCTCAGATGCTGCGTACCGGAAACAACCTTTTACCTTTCAGCTACACTTTTGGAGACGCTCT 3487
QY 421 PheHisSerSerTrrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
DB 3488 TTCCACAGCAGCTACGCTCACAGCAGAGTCTGGACCGCTCTCATGAATCTCTCATCGAC 3547
QY 441 GlnTrrLeuTrrTrrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAsp 460
DB 3548 CAGTACCTGTATTACTTTGAGCAGACAACAACACTCCAAGTGGAAACCCACACGAGCTCAAGG 3607
QY 461 LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrrLeuPro 480
DB 3608 CTTCAAGTTTCTCAGCGCGGAGCGAGTGAACATTTGGGACCGAGCTTAGGAATCGGCTTCT 3667
QY 481 GlyProCysTrrArgGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsn 500
DB 3668 GGACCTCTTACCAGCAGCAGAGTAGTCAAAAGACATCTCGCGGATAACAACAACAGTGAA 3727
QY 501 PheThrTrrPrrGlyAlaSerLysTrrAsnLeuAsnGlyArgGluSerIleIleAsnPro 520
DB 3728 TACTGTGGAGCTGGAGCTTCAACAGTACCACCTCAATGGCAGAGACTCTCTGGTGAATCCG 3787
QY 521 GlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyVal 540
DB 3788 GCGCGCGCGCATGGCAGGACGACAGGACGATGAGAAAAGTTTTCCTCAGAGCGGGGT 3847
QY 541 MetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIle 560
DB 3848 CTCATCTTTGGAGCAAGGCTCAGAGAAAACAATGTGGACATGTAAGAAAGTCTCTCAGAGCGGGGT 3907
QY 561 ThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrVal 580
DB 3908 ACAGACGAGAGGAATCAGGACAACAATCCCGTGGCTACGAGACGATATGGTTCTGTGA 3967

```


Db 2683 AAGCGGGCCAGCAGCCTCGAAGAAAAAGATTGAAATTTGGTCCAGACTGGAGCAGCAGAC 2742
Qy
181 SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro 200
Db 2743 TCAGTACCTGACCCCGAGCCTCTCGGACAGCCACCCAGCAGCCCTCTGGTCTGGAACT 2802
Qy 201 ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnGlnGluAlaAlaAspGly 220
Db 2803 AATACGATGGCTACAGCAGTGGCGACCACTAATGGCAGACATTAACGAGGGCGCCGACGGA 2862
Qy 221 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValle 240
Db 2863 GTGGGTAAATTCGCGGAAATTTGGCAATTCGATTCACATGGATGGCGCAGACAGTCATC 2922
Qy 241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrGlyGlnIle 260
Db 2923 ACCACGAGCAGCCGGAACCTTGGGGCCCTGCCACCCTACACACACCCTCTACAAACAAAT 2982
Qy 261 SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrp 280
Db 2983 TCCAGCCAAATCA--GGAGCTCGAAGCAGCAATCACTACTTTGGCTACAGCAGCCCTTGG 3039
Qy 281 GlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu 300
Db 3040 GGGTATTTTACCTTCAACAGATTCCACTGCGCACTTTTCCACAGCTGACTGGCAAGACTC 3099
Qy 301 IleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGln 320
Db 3100 ATCAACAACAACCTGGGGATTTCCGACCCAGAGACTCAACTTCAAGCTCTTTAATCAATCAA 3159
Qy 321 ValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThr 340
Db 3160 GTCAAGAGGTCACGCAGCAATGACGGTACGACGACGATTCGCCAATAACCTTACCAGCAG 3219
Qy 341 ValGlnValPheSerAspSerGlyLysThrGlnLeuProTyrValLeuGlySerAlaHisGln 360
Db 3220 GTTCAGGTGTTTACTGACTCGGAGTACCACTCCGCTCCGCTCCGCTCGGCGCATCAA 3279
Qy 361 GlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr 380
Db 3280 GGATGCTCCCGCGTTCGCCAGAGAGCTTTTATGTCGCCAGATGATGATACCTCACCC 3339
Qy 381 LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlyLysThrPhePro 400
Db 3340 CTGAACAACGGGAGTCAGGCAGTAGGAGCGCTCTTCAATTTTACTGCTGGAGTACTTTCT 3399
Qy 401 SerGlnMetLeuArgThrGlyAsnAsnAsnPheThrPheSerTyrThrPheGluGluValPro 420
Db 3400 TCTCAGATGCTGGGTACCGGAAACAACTTTACCTTCAGCTACACATTTTGGAGGACGTCTCT 3459
Qy 421 PheHisSerTyrTrpAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
Db 3460 TTCACAGCAGCTTACGCTCACGCCAGAGTCTGGACCGTCTCATGATCTCTCTCATCGAC 3519
Qy 441 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAsp 460
Db 3520 CAGTACCTGTATTACTTGACGAGCAACAAACACTCCAAAGTGGAAACCAACCAAGCTCAAGG 3579
Qy 461 LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuPro 480
Db 3580 CTTTCAGTTTTTCTCAGCGCGGAGGAGTGCATTTCCGGACCACTGTAGGAACCTGGCTTCT 3639
Qy 481 GlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsn 500
Db 3640 GGACCTTGTTACCGCCAGCGAGGATATCAAGACATCTCGGATTAACAACAACAGTGA 3699
Qy 501 PheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleLeuAsnPro 520
Db 3700 TACTCGTGGAGCTGGAGCTACCAAGTACCACTCAATGGCAGAGACTCTCTGTGTAATCCG 3759
Qy 521 GlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyVal 540
Db 3760 GGCCCGGCCCATGGCAAGCCCAAGGACGATGAAGAAAGTTTTTTCTCTCAGAGCGGGTT 3819

Qy 541 MetIlePheGlyLysArgSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIle 560
Db 3820 CTCATCTTTGGAGACGAGCGCTCAGAGAAAACAATGTGGACATTTGAAAGGTCTATGATT 3879
Qy 561 ThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrVal 580
Db 3880 ACAGACGAGAGGAAATCAGGACCAACCAATCCCGTGGCTACGGAGCAGTGTGTTCTGTA 3939
Qy 581 AlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGly 600
Db 3940 TCTACCAACCTCCAGAGAGGCAACAGACAGCAGTACCCGAGATGTCAACACACAAGGC 3999
Qy 601 AlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAla 620
Db 4000 GTTCTTCCAGCATGGTCTGCGAGGACAGAGATGTGTACCTTCAGGGGCCCATCTGGCA 4059
Qy 621 LysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeu 640
Db 4060 AAGATTCCACACACGAGCGACATTTTCCACCCCTCTCCCTCATGGGTGGATTCGGACTT 4119
Qy 641 LysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAla 660
Db 4120 AAACACCTCTCTCCAGATTTCTCATCAAGAACACCCCGGTACCTGCCAATCTCTCGACC 4179
Qy 661 GluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSer 680
Db 4180 ACCTTCAGTCGGCAAGTTTGTCTTCTCATCACAGTACTCCACGGACAGGTTCAGC 4239
Qy 681 ValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTyrAsnProGluValGln 700
Db 4240 GTGGAGATCGAGTGGGAGCTGCAGAAAGAAACAGCAACGCTGGAATCCCGAAATTCAG 4299
Qy 701 TyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeu 720
Db 4300 TACATCTTCAACTACCAAGTCTGTTAATGTGGACTTTTACTGTGACACTAATGGCGTG 4359
Qy 721 TyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 4360 TATTCAGAGCTCGCCCCATTGGCACCAGATACCTGACTCGTAATCTG 4407

RESULT 13
AAT09008
ID AAT09008 standard; DNA; 4680 BP.
XX AAT09008;
AC AAT09008;
XX 05-JUL-1996 (first entry)
DT
DE Wild-type adeno-associated virus 2 genome.
XX adeno-associated virus 2; AAV2; inverted terminal repeat; viral vector;
KW DNA delivery; immunodeficiency virus protein; immunity; human; simian;
KW neurological disorder; therapy; ss.
XX Adeno-associated virus.
OS
XX W09534670-A2.
PN
XX 21-DEC-1995.
PD
XX 06-JUN-1995; 95WO-US07178.
PF
XX 06-JUN-1994; 94US-0254358.
PR
XX (CHIL-) CHILDRENS HOSPITAL INC.
PA
XX Johnson PR;
PI
XX WPI; 1996-049697/05.
DR
XX Recombinant adeno-associated virus genome contg. protein encoding
PT DNA - flanked by inverted terminal repeats, for use in vaccines or

for treatment of neuro-degenerative disease

Example 1; Page 27-29; 44pp; English.

The present sequence is that of the wild-type adeno-associated virus (AAV) 2, a replication-deficient parvovirus. Cis-acting sequences directing viral DNA replication (ori), encapsidation/packaging (pkag) and host cell chromosome integration (int) are contained within the ITRs (inverted terminal repeats of 145 nucleotides). When AAV infects a human cell, the viral genome integrates into chromosome 19 resulting in latent infection of the cell. Prodn. of infectious virus does not occur unless the cell is infected with a helper virus (e.g., adenovirus or herpes- virus). AAV possesses unique features that make it attractive as a vector for delivering foreign DNA to cells. A vector including a recombinant AAV genome contg. a simian immunodeficiency virus (SIV) rev and envelop (gp160) gene cassette was constructed from an existing plasmid designated psub201. The plasmid contains a modified wild-type AAV2 genome. Specifically, two XbaI sites were added via linker addition at sequence positions 190 and 4484. These sites are internal to 191 bp ITRs which include the 145 bp ITRs of the AAV genome. The insertion of these sites allows the complete removal of the internal 4.3 kb fragment contg. the AAV rep-cap genes upon XbaI digestion of the plasmid. (See also AAT09009-10).

Sequence 4680 BP; 1198 A; 1265 C; 1254 G; 963 T; 0 other;

Alignment Scores:

| | | | |
|------------------------|-----------|---------------|------|
| Pred. No.: | 8,619-237 | Length: | 4680 |
| Argument Scores: | | Matches: | 609 |
| Score: | 3360.50 | Conservative: | 502 |
| Percent Similarity: | 89.4% | Mismatches: | 76 |
| Best Local Similarity: | 82.63% | Indels: | 2 |
| Query Match: | 84.24% | Gaps: | 1 |
| DB: | 17 | | |

US-09-807-802A-13 (1-736) X AAT09008 (1-4680)

| | | | |
|----|------|---|------|
| Qy | 1 | MetAlaAlaAepGlyTyrLeuProAspTrpLeuGluAepAenLeuSerGluGlyIleArg | 20 |
| Db | 2203 | ATGGCTCGCATGGTTATCTTCCAGATTGGCTCGAGGACACTCTCTCTGAAGGAATAAGA | 2262 |
| Qy | 21 | GluTrpTrpAspLeuLysProGlyAlaProLysProLysAlaAenGlnGlnLysGlnAsp | 40 |
| Db | 2263 | CAAGTGGTGAAGCTCAAACTTGGCGCCACCAACCAAGCCGCGAGAGCGGCATAAGGAC | 2322 |
| Qy | 41 | AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAenGlyLeuAsp | 60 |
| Db | 2323 | GACAGCAGGGGTCTTGTGCTCTCTGGGTACAAAGTACCTCGGACCCCTTCAACGAGCTCGAC | 2382 |
| Qy | 61 | LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp | 80 |
| Db | 2383 | AAGCGAGAGCCGGTCAACGAGCGAGACGCGCGCGCCCTCGAGCACGACAAAGCCCTACGAC | 2442 |
| Qy | 81 | GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAAspAlaGluPhe | 100 |
| Db | 2443 | CGGCAGCTCGACAGCGGAGACAAACCCGTACCTCAAGTACAAACCAACGCGGACGCGGAGTTT | 2502 |
| Qy | 101 | GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln | 120 |
| Db | 2503 | CAGAGCGCCCTTAAAGAAGATACGTCTTTTGGGGCAACCTCGGAGCAGCGAGTCTTTCAG | 2562 |
| Qy | 121 | AlaLysLysArgValLeuGluProLeuGlyLeuValGluGlyAlaLysThrAlaPro | 140 |
| Db | 2563 | CGCAAAAAGAGGGTTCTTGAACCTCTGGGCCTGGTTGAGGAACCTGTTAAGACGGCTCCG | 2622 |
| Qy | 141 | GlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSerGlyIleGly | 160 |
| Db | 2623 | GGAAAAAAGAGCGCGGTAGAGCACTCTCTGTGAGCGCAGACTCTCTCTCGGGAAACCGGA | 2682 |
| Qy | 161 | LysThrGlyGlnGlnProAlaLysLysLysArgLeuAenPheGlyGlnThrGlyAspSerGlu | 180 |
| Db | 2683 | AAGCGGGCCACAGCCCTGCAGAAAAAAGATTGTAATTTTGTCTAGACTGTGGAGACGGACG | 2742 |
| Qy | 181 | SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro | 200 |


```
Db 2683 AAGCGGCGCCAGCAGCTGCAAGAAAGATTGAATTTGGTCAGACTGGAGAGCGCAGAC 2742
Qy 161 SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro 200
Db 2743 TCAGTACCTGACCCCGCCCTCTCGGACAGCCAGCAGCCCTCTGGTCTGGGAAT 2802
Qy 201 ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGly 220
Db 2803 AATACGATGGCTACAGGAGTGGCGCACCATTGGCAGACATTAACGAGGGCGCCGCGGA 2862
Qy 221 ValGlyAenAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValile 240
Db 2863 GTGGGTAAATCTCTCGGAAATTTGGCATTTGCCATTCCACATGGATGGCGAGCAGAGTCATC 2922
Qy 241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnile 260
Db 2923 ACCACGACCCCGAACCTGGGCCCTGCCACCTACAAACACCCTCTCAACAACAAAT 2982
Qy 261 SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrp 280
Db 2983 TCCAGCCCAATCA--GGAGCCTCGAACGACAAATCACTACTTTGGCTACAGCACCCTTTGG 3039
Qy 281 GlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu 300
Db 3040 GGGTATTTTGACTTCAACAGATTCCACTGCCACTTTTCCACCACGTGACTGGCAAGACTC 3099
Qy 301 IleAsnAsnAntTrpGlyPheArgProLysArgLeuAsnPheLeuPheAsnIleGln 320
Db 3100 ATCAACAACAACCTGGGGATTCCGACCCCAAGAGACTCAACTTCAAGCTCTTTAACAATCCA 3159
Qy 321 VallysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThr 340
Db 3160 GTCAAGAGGTCAACGACAAATGACGGTACGACGACGATTGGCAATAAACCCTTACCAGCAGC 3219
Qy 341 ValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGln 360
Db 3220 GTTCAAGGTGTTTACTGACTCGAGTACCGAGTCCCGTACGCTCCGCTCGGCTCGGCGCATCAA 3279
Qy 361 GlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr 380
Db 3280 GGATGCTCTCCCGCGTTCACGACAGAGCTTTCATGGTGCCACAGATAGGATACCTCACC 3339
Qy 381 LeuAsnAsnGlySerGlnAlaValGlyArgSerPheTyrCysLeuGluTyrPhePro 400
Db 3340 CTGAACAACAGGAGTCAGCAGTAGGAGCGCTCTTCAATTTACTGCTGGAGTACTTTCCT 3399
Qy 401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 420
Db 3400 TCTCAGATGCTGCTGACCGGAAACAACTTTACCTTCAGCTACACTTTTGGAGGACGTTCTCT 3459
Qy 421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
Db 3460 TTCCACAGCAGCTAGCTCACAGCCAGAGTCTGGACCGCTCATGAATCCTCTCATCGAC 3519
Qy 441 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAsp 460
Db 3520 CAGTACCTGTATTACTTGACGAGAAACAAACACTCCAAGTGGAAACCCACCAGCAGTCAAGG 3579
Qy 461 LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuPro 480
Db 3580 TTTCAGTCTTCTCAGCGCGGAGTGCATTCGGGCCAGTCTAGGAACCTGGCTTCTCT 3639
Qy 481 GlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsn 500
Db 3640 GGACCTCTGTACCGCCAGCAGCGAGTATCAAGAGCATCTGCGGATAACAACAACAGTGAA 3699
Qy 501 PheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleLeuAsnPro 520
Db 3700 TACTCGTGGACTGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTGTGTAATCCG 3759
Qy 521 GlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyVal 540
```

```
Db 3760 GGGCCCGCCATGGCAAGCCCAAGGACGATGAAGAAAGATTTTTTCTCAGAGCGGGTT 3819
Qy 541 MetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIle 560
Db 3820 CTATCTTTGGGAAGCAGGCTCAGAGAAACAATGTGAACATTGAAAGGTCATGATT 3879
Qy 561 ThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrVal 580
Db 3880 ACAGACGAGGAGAAATCGGAACAACCAATCCCGTGTCTACGGAGCAGTAGTGGTTCTGTA 3939
Qy 581 AlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGly 600
Db 3940 TCTACCAACTCCAGAGAGCAACAGACGATACCCGAGATGTCAACAACAAGGC 3999
Qy 601 AlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAla 620
Db 4000 GTTCTTCAGGCATGGTCTGGCAGGACAGAGATGTGTACCTTCAGGGGCCCATCTGGGCA 4059
Qy 621 LysIleProHisThrAspGlyHisPheHisPheSerProLeuMetGlyGlyPheGlyLeu 640
Db 4060 AAGATTCCACACGAGCGGACGATTTTTCAGCCCTCTCCCTCATGGGTGGATTGGACTT 4119
Qy 641 LysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsnProAla 660
Db 4120 AACACCCCTCTCCACAGATTCATCAAGAACAACCCCGGTACCTGCGAATCTCTCGACC 4179
Qy 661 GluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGln-ValSe 680
Db 4180 ACCTTCAGTGGCGAAAGTTTGTCTTCTCATCACAGACTTCCACGGGACACGGTCAG 4239
Qy 680 rValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValG 700
Db 4240 CGTGGAGATCGAGTGGGAGCTGCAGAAAGAAACAGCAACCGCTGGAATCCCGAAATTC 4299
Qy 700 nTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnGlyLe 720
Db 4300 GTACACTTCAACTACAAACAGTCTGTAATCGTGGACTTACCGTGGATATAATGGCT 4359
Qy 720 uTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 4360 GTATTACAGGCTCGCCCCATTTGGCACCAGATACCTGACTCGTAATCTG 4408

RESULT 15
AAH41481
ID AAH41481 standard; DNA; 4675 BP.
XX AC AAH41481;
XX AC AAH41481;
XX DT 23-AUG-2001 (first entry)
XX DE Adeno-associated virus (AAV) plasmid pAV1 AvaiI nucleotide sequence.
XX KW Adeno-associated virus; AAV; gene transfer; adenovirus; chromosome 19;
XX KW medicine; cell engineering; genetic engineering; treatment;
XX KW embryological engineering; ds.
XX OS Adeno-associated virus.
XX OS WO200132899-A1.
XX PN 10-MAY-2001.
XX PF 23-OCT-2000; 2000WO-JP07373.
XX PR 29-OCT-1999; 99JP-0308839.
XX PR (TAKI ) TAKARA SHUZO CO LTD.
XX PA Ueno T, Matsumura H, Tanaka K, Iwasaki T, Ueno M, Fujinaga K;
XX PI Asada K, Kato I;
XX DR WPI; 2001-316450/33.
XX
```

PT Gene transfer method for highly efficient introduction of foreign genes
 PT to cells in humans particularly by integration specifically into AAVS1
 PT site of chromosome 19 for transformation, applicable e.g. in disease
 PT treatment -

XX Example 1; Page 28-31; 38pp; Japanese.

XX The present invention describes a gene transfer method for highly
 CC efficient introduction of foreign genes to cells in humans particularly
 CC by integration specifically into AAVS1 site of chromosome 19 for
 CC transformation, applicable e.g. in disease treatment. The method
 CC comprises transferring into the cells (with the use of an adenovirus
 CC vector) a nucleic acid which has a sequence provided with adeno-
 CC associated virus (AAV)-originated inverted terminal repeats (ITRs) in
 CC both sides of the target foreign gene to be transferred, a second nucleic
 CC acid which has an AAV-originated rep gene and a promoter for expressing
 CC this gene and carries a stuffer sequence inserted into it sandwiched in
 CC two recombinase-recognition sequences and located between the rep gene
 CC and promoter and expressing the Rep protein under the action of the
 CC recombinase in the cells obtained previously to integrate the target
 CC foreign gene into the chromosomal DNA. The method is for transferring a
 CC foreign gene into cells particularly in human, especially by integration
 CC specifically into AAVS1 site of chromosome 19 for transformation, which
 CC is applicable in medicine, as well as cell, genetic and embryological
 CC engineering e.g. in disease treatment. The method is highly efficient,
 CC with use of adenovirus vectors, nucleic acids and other sequences
 CC including rep genes and promoters, by expressing rep proteins to.
 CC integrate target foreign gene into chromosomal DNA through action of
 CC recombinase. The present sequence represents an AAV plasmid pAV1 Avall
 CC nucleotide sequence which is used in an example from the present
 CC invention.

XX SQ Sequence 4675 BP; 1198 A; 1262 C; 1251 G; 964 T; 0 other;

Alignment Scores:

Precl. No.: 5.3e-234 Length: 4675
 Score: 3322.50 Matches: 607
 Percent Similarity: 89.02% Conservative: 50
 Best Local Similarity: 82.25% Mismatches: 76
 Query Match: 83.29% Indels: 5
 DB: 22 Gaps: 2

US-09-807-802A-13 (1-736) x AAH41481 (1-4675)

QY 1 MetAlaAlaSpGlyTyrLeuProAspTrpLeuGluAspAsnLeuSerGluGlyIleArg 20
 Db 2203 ATGGCTGCCGATGGTTATCTTCAGATTGGCTCGAGGACACTCTCTCTGAAGGAATAGA 2262
 QY 21 GluTrpTrpAspLeuLysProGlyAlaProLysProLysAlaAsnGlnGlnLysGlnAsp 40
 Db 2263 CAGTGTGGAGTCAAACTCGCCACCACCAACCCAGCCGAGCGGCATAGGAC 2322
 QY 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
 Db 2323 GACAGCAGGGGCTTGTGCTCTCTGGGTCAAGTACCTCGACCCCTTCAACGACATCGAC 2382
 QY 61 LysGlyGluProValAsnAlaAlaAspAlaAlaAlaLeuCluHis-AspIysAlaTyrAs 80
 Db 2383 ARGGGAGAGCCGGTCAACGAGGAGAGCGCCGCGCCCTCGAGCAGCTACAAAGCCCTACGA 2442
 QY 80 pGlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPh 100
 Db 2443 CCGGCAGCTCGACAGCGGAGCAACCCGTACTCTCAAGTACAACACCGCCGCGGAGTT 2502
 QY 100 eGlnGluArgLeuGlnGluAspThrSerPheGlyAsnLeuGlyArgAlaValPheG1 120
 Db 2503 TCAGGAGCGGCTTAAAGAGATACGTCTTTTGGGGCAACTCTCGGCGAGCAGCTCTTCCA 2562
 QY 120 nAlaLysLysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaPr 140
 Db 2563 GGGGAAAAGAGGGTCTTGAACCTCTGGGCCCTGGTTGAGGAACCTGTTAAGACGGCTCC 2622
 QY 140 oGlyLysArgProValGluGlnSerProGlnGluProAspSerSerSerGlyIleG1 160

Db 2623 GGGAAAAGAGGCGCGTAGAGCACTCTCTGTGGAGCCAGACTCTCTCTCGGAGCCGG 2682
 QY 160 YLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerG1 180
 Db 2683 AAAGCGGCGCCAGCAGCCTGCAAGAAAAGATTGAATTTTGGTCAGACTGGAGACGCAGA 2742
 QY 180 uSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPr 200
 Db 2743 CTCAGTACCTGACCCCGAGCCTCTCGGACACCCAGCCAGCAGCCCTCTGTGTGGGAAC 2802
 QY 200 oThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspG1 220
 Db 2803 TAATACGATGGCTACAGGAGCTGGCGCACCAATGGCAGACAATAACGAGGCGCGGACGG 2862
 QY 220 yValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgVal11 240
 Db 2863 AGTGGTAAATTCCTCGGAAATTCGATTGGCATTTGCCATGTGGATGGGCGCAGAGTGTAT 2922
 QY 240 eThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGln1 260
 Db 2923 CACCACAGCAGCACCAGAACCTGGGCGCTGCCACCTTACAAACAACCACTCTTACAAACAAT 2982
 QY 260 eSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTr 280
 Db 2983 TTCCAGCCCAATCA--GGAGCCTCGACGACAATCACTACTTTGGCTACAGCACCCTTG 3039
 QY 280 pGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLe 300
 Db 3040 GGGGTATTTTACATTTCAACAGATTCCACTGCCACTTTTCCACACGATGATGGCAAAAGACT 3099
 QY 300 uIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleG1 320
 Db 3100 CATCAACAACAACCTGGGGATTCGACCCCAAGAGACTCAACTTCAAGCTCTTTACATTCA 3159
 QY 320 nValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerTh 340
 Db 3160 AGTCAACAGAGTCCAGCAGAAATGACGGTACGACGAGCATTTGCCAATAACCTTACCAGCAC 3219
 QY 340 rValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisG1 360
 Db 3220 GGTTCAGGTGTTTACTGACTCGGAGTACCACTCCCGTACGCTCCGCTCGGCGCATCA 3279
 QY 360 nGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuTh 380
 Db 3280 AGGATGCTCTCCGCGCTTCCAGCAGAGAGCTTTCATGTGTGCGCACAGTATGGATACCTCAC 3339
 QY 380 rIleAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePr 400
 Db 3340 CCTGAACAACGGGAGTCAGGAGTACGAGCGCTCTTCACTTTTACTGCTGGAGTACTTTCC 3399
 QY 400 oSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPr 420
 Db 3400 TTCTCAGATGTGCTGCTACCGGAAACAACCTTACCTTCAGCTACACTTTTGGAGAGCTTCC 3459
 QY 420 oPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAs 440
 Db 3460 TTTCCACAGCAGCTACGCTCACAGCCAGAGTCTGGACCGTCTCTCATGAATCTCTCATCGA 3519
 QY 440 pGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAs 460
 Db 3520 CCAGTACCTGTATTACTTGGAGCAACAACAACCTCAAGTGGAAACCAACCCAGCAGCTCAAG 3579
 QY 460 pIleuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuPr 480
 Db 3580 GCTTCAGTTTCTTCAGCGCGGAGCGAGTGACACTTCGGGACCACTCTAGGAACTTGGCTTCC 3639
 QY 480 oGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAs 500
 Db 3640 TGGACCTGTATTACCGCAGCAGCAGTATCAAAAGACATCTGGGAGATACCAACAACAGTGA 3699
 QY 500 nPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnPr 520


```
Db 3700 ATACTCTGTGACTGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTGTGTGAATCC 3759
Qy 520 oGlyThrAlaMetAlaSerHisLysAspGluAspLysPhePheProMetSerGlyVa 540
Db 3760 G-----GCCATGGCAAGCCACAGGACGATGAAGAAAGTTTTTCTTCAGAGCGGGT 3813
Qy 540 lMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIl 560
Db 3814 TCTCATCTTTGGGAAGCAAGCTCAGAGAAACAAATGTGAACATTGAAAAAGGTCATGAT 3873
Qy 560 eThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrVa 580
Db 3874 TACAGACGAGAGGAATTCGAACACCAATCCCGTGGCTACGGAGCAGTATGGTCTGT 3933
Qy 580 lAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGl 600
Db 3934 ATCTACCAACCTCCAGAGAGGCAACAGACAGCAGCTACCGCAGATGTCAACACACAAGG 3993
Qy 600 yAlaLeuProGlyMetValTrpGlnAspArgAspValTyLeuGlnGlyProIleTrpAl 620
Db 3994 CGTTCTTCCAGGCATGGCTGGCAGSACAGAGATGTGTACCTTCAGGGGCCCATCTGGGC 4053
Qy 620 aLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLe 640
Db 4054 AAAGATTCCACACACGACGACGACATTTTCAACCCCTCTCCCTCATGGGTGGATCGGACT 4113
Qy 640 uLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAl 660
Db 4114 TAAACACCTCTCTCCACAGATTCTCATCAAGAACACCCCGGTACCTCGGAATCCTCGAC 4173
Qy 660 aGluPheSerAlaThrLysPheAlaSerPheIleThrGlnThrSerThrGlyGln-ValS 680
Db 4174 CACCTTCAGTGGCGCAAGTTTGCTTCCTTCATCACACAGTACTCCACGGGACACGGTCA 4233
Qy 680 erValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValG 700
Db 4234 GCGTGGAGATTCAGTGGGAGCTGCAGAGGAAACAGCAACGCTGGAAATCCCGAAATTC 4293
Qy 700 lNtyrThrSerAsnTyralAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyL 720
Db 4294 AGTACACTTCCAACCTACAACAAGTCTGTTAATCGTGGACTTACCGTGATATAATGGCG 4353
Qy 720 eutyThrGluProArgProIleGlyThrArgTyLeuThrArgProLeu 736
Db 4354 TGTATTACAGGCCTGGCCCCCATGGCACCATACCTGACTCGTAATCTG 4403
```

Search completed: February 14, 2004, 17:19:59
Job time : 583.843 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 14, 2004, 16:11:58 : Search time 108.687 Seconds
(without alignments)
2988.931 Million cell updates/sec

Title: US-09-807-802A-13
Perfect score: 3989
Sequence: 1 MAADGYLPDWNLEDSGIR.....NNGLYTEPRIGTYRLTRPL 736

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Xgapop 10.0, Xgapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 59978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_epool/US09807802/runat_11022004_175609_15955/app_query.fasta_1.2389
-DB=Issued Patents NA -QWMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTWT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09807802@cgn2_1_131 @runat_11022004_175609_15955 -NCPU=6 -ICPU=3
-NO_WMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------------|
| 1 | 3439 | 86.2 | 8179 | 4 | US-09-438-268-5 |
| 2 | 3402.5 | 85.3 | 4072 | 4 | Sequence 5, Appli |
| 3 | 3402.5 | 85.3 | 7557 | 4 | Sequence 4, Appli |
| 4 | 3402.5 | 85.3 | 8698 | 4 | Sequence 3, Appli |
| 5 | 3360.5 | 84.2 | 4680 | 1 | Sequence 2, Appli |
| 6 | 3360.5 | 84.2 | 4680 | 1 | Sequence 1, Appli |
| 7 | 3360.5 | 84.2 | 4680 | 1 | Sequence 1, Appli |
| 8 | 3360.5 | 84.2 | 4680 | 5 | Sequence 1, Appli |
| 9 | 3038.5 | 75.2 | 8151 | 4 | Sequence 2, Appli |
| 10 | 2486.5 | 62.3 | 2208 | 4 | Sequence 5, Appli |
| c 11 | 2486.5 | 62.3 | 7214 | 4 | Sequence 1, Appli |
| 12 | 2481.5 | 62.2 | 4767 | 4 | Sequence 1, Appli |

| | | | | | | |
|------|--------|------|---------|---|----------------------|--------------------|
| 13 | 1830.5 | 45.9 | 1800 | 4 | US-09-532-594B-17 | Sequence 17, Appli |
| 14 | 1690.5 | 42.4 | 1617 | 4 | US-09-532-594B-19 | Sequence 19, Appli |
| 15 | 1410 | 35.3 | 2271 | 4 | US-09-438-268-3 | Sequence 3, Appli |
| 16 | 439 | 11.0 | 5049 | 1 | US-08-336-345-1 | Sequence 1, Appli |
| 17 | 439 | 11.0 | 5049 | 1 | US-08-336-345-2 | Sequence 2, Appli |
| 18 | 439 | 11.0 | 5049 | 2 | US-08-647-655-1 | Sequence 1, Appli |
| 19 | 439 | 11.0 | 5049 | 2 | US-08-647-655-2 | Sequence 2, Appli |
| 20 | 263 | 6.6 | 2062 | 6 | 5223424-3 | Patent No. 5223424 |
| 21 | 247 | 6.2 | 1740 | 1 | US-07-969-213-1 | Sequence 1, Appli |
| 22 | 220 | 5.5 | 1820 | 6 | 5223424-12 | Patent No. 5223424 |
| 23 | 203 | 5.1 | 2254 | 3 | US-08-552-369-1 | Sequence 1, Appli |
| 24 | 199.5 | 5.0 | 5910 | 1 | US-08-195-814-1 | Sequence 1, Appli |
| 25 | 192 | 4.8 | 1752 | 3 | US-09-022-949-1 | Sequence 1, Appli |
| 26 | 184 | 4.6 | 1755 | 2 | US-08-317-785-1 | Sequence 83, Appli |
| 27 | 132 | 3.3 | 3906 | 4 | US-09-996-243-83 | Sequence 13774, A |
| c 28 | 131 | 3.3 | 4188 | 4 | US-09-252-991A-13774 | Sequence 13656, A |
| 29 | 131 | 3.3 | 8211 | 4 | US-09-452-991A-13656 | Sequence 8, Appli |
| c 30 | 130 | 3.3 | 1896 | 1 | US-08-198-446B-8 | Sequence 8, Appli |
| c 31 | 130 | 3.3 | 1896 | 2 | US-08-870-693-8 | Sequence 930, App |
| c 32 | 127.5 | 3.2 | 4563 | 4 | US-09-252-991A-930 | Sequence 1, Appli |
| c 33 | 126.5 | 3.2 | 8438 | 1 | US-07-945-283-1 | Sequence 2, Appli |
| c 34 | 126 | 3.2 | 4403765 | 3 | US-09-103-840A-2 | Sequence 1, Appli |
| c 35 | 126 | 3.2 | 4411529 | 3 | US-09-103-840A-1 | Sequence 1, Appli |
| c 36 | 125 | 3.1 | 4214 | 4 | US-09-122-135-1 | Sequence 1, Appli |
| c 37 | 123.5 | 3.1 | 3451 | 4 | US-09-811-286-1 | Sequence 11853, A |
| 38 | 123 | 3.1 | 2997 | 4 | US-09-252-991A-11853 | Sequence 1085, Ap |
| 39 | 122 | 3.1 | 2817 | 4 | US-09-620-312D-1085 | Sequence 1282, Ap |
| 40 | 121 | 3.0 | 2305 | 4 | US-09-016-434-1282 | Sequence 1, Appli |
| 41 | 119.5 | 3.0 | 4319 | 5 | PCT-US95-10661A-1 | Sequence 1, Appli |
| 42 | 119.5 | 3.0 | 3288 | 4 | US-09-107-532A-201 | Sequence 201, App |
| 43 | 118.5 | 3.0 | 3546 | 3 | US-08-872-757-3 | Sequence 3, Appli |
| 44 | 118.5 | 3.0 | 3546 | 4 | US-09-850-048A-3 | Sequence 3, Appli |
| 45 | 118.5 | 3.0 | 3546 | 4 | US-09-850-048A-3 | Sequence 3, Appli |

ALIGNMENTS

RESULT 1

US-09-438-268-5
; Sequence 5, Application US/09438268
; Patent No. 6491907
; GENERAL INFORMATION:
; APPLICANT: Rabinowitz, Joseph E.
; APPLICANT: Samulski, Richard J
; APPLICANT: Xiao, Weidong
; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 5470-186
; CURRENT APPLICATION NUMBER: US/09/438,268
; CURRENT FILING DATE: 1999-11-10
; EARLIER APPLICATION NUMBER: 60/107,840
; EARLIER FILING DATE: 1998-11-10
; EARLIER APPLICATION NUMBER: 60/123,651
; EARLIER FILING DATE: 1999-03-10
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 8179
; TYPE: DNA
; ORGANISM: Virus
US-09-438-268-5

Alignment Scores:
Pred No.: 0
Score: 3439.00
Percent Similarity: 90.50%
Best Local Similarity: 84.94%
Query Match: 86.21%
DB: Gaps: 2
Matches: 626
Conservative: 41
Mismatches: 68
Indels: 2

US-09-807-802A-13 (1-736) x US-09-438-268-5 (1-8179)

Qy 1 MetAlaAlaepGlyTyrLeuProAspTrpLeuGluAspAsnLeuSerGluGlyLeArg 20

Db 2133 ATGGTCGCGATGTTATCTTCCAGATTGGCTCGAGGACACTCTCTCTGGAAGGAATAAGA 2192
Qy 21 GlnTrrpTrrpAspLeuLysProGlyAlaProLysProLysAlaAsnGlnGlnLysGlnAsp 40
Db 2193 CAGTGGTGAAGCTCAAACTGGCCACACACCAACCAAGCCGCGAGCGGCATAGGAC 2252
Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
Db 2253 GACAGCAGGGTCTTGTCTTCTGGGTACAAGTACCTCGGACCCTTCAACGGACTCGAC 2312
Qy 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
Db 2313 AAGCGAGAGCGGTCAACGAGGACAGACCGCGCGCTCGAGCAGCAAAAGCCCTACGAC 2372
Qy 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
Db 2373 CGCGAGCTCGACAGCGGAGACAAACCGGTACCTCAAGTACACCAAGCCGCGAGGAGTTT 2432
Qy 101 GlnGluArgLeuGlnGlnAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln 120
Db 2433 CAGGAGCGCTTAAAGAGATACGTCTTTTGGGGCAACCTCGGACGACGAGTCTTCCAG 2492
Qy 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGlnGluGlyAlaLysThrAlaPro 140
Db 2493 GCGAAAAGAGGGTCTTGAACCTCTGGGCCCTGGTTGAGGAACCTGTTAAGACGGCTCG 2552
Qy 141 GlyLysLysArgProValGlnGlnSerProGlnGluProAspSerSerSertGlyIleGly 160
Db 2553 GGAAGAAAGAGCGGTAGAGCACCTCTCTGTGGAGCCAGACTCTCTCTCGGGAACCGGA 2612
Qy 161 LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
Db 2613 AAGCGGGCGCAGCAGCTTCAAGAAAGAAAGATTGGTATTTGGTTCAGACTGGAGACGAC 2672
Qy 181 SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro 200
Db 2673 TCAGTACCTTGACCCCGCCCTCTCGGACAGCCACACAGCCCTCTGTGTGGGAAT 2732
Qy 201 ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnGlnGlyAlaAspGly 220
Db 2733 AATACGATGGCTACAGCAGTGGCGCACCAATGGCAGACAATAACGAGGGCGCCGACGGA 2792
Qy 221 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIle 240
Db 2793 GTGGGTAAATTCCTCCGAAATTTGGCATTTCCATGATGGATGGCGACAGAGTCAATC 2852
Qy 241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIle 260
Db 2853 ACCACGAGCACCCGAACTGGGCCCTGGCCACCTACACACACCACTCTACAAACAAATT 2912
Qy 261 SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrp 280
Db 2913 TCCAGCCAAATCA--GGAGCTCGAACGACCAATCACTACTTTGGCTACAGCACCCCTGG 2969
Qy 281 GlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu 300
Db 2970 GGGTATTTTGACTTCAACAGATTCCACTGCCATTTTCAACCGTACGTCGGCAAGACTC 3029
Qy 301 IleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGln 320
Db 3030 ATCAACACAACTCGGGGATTCGACCCCAAGAGACTCACTTCAAGTCTTTAATCAATCA 3089
Qy 321 ValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThr 340
Db 3090 GTCAAGAGGTACCGCAGAAATGACGCTAGCAGCAGCAATTAACCTTACCAACGACG 3149
Qy 341 ValGlnValPheSerAspSerGlyTyrGlnLeuProTyrValLeuGlySerAlaHisGln 360
Db 3150 GTTCAGGTGTACTCACTCGGAGTACCACTCCCGTACGTCTCGGGTCGGCGCACCAA 3209
Qy 361 GlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr 380

Db 3210 GGCTGTCTCCCGCGGTTCAGCGGACGCTTTTCATGGTCCCTCAGTATGATACCTCACC 3269
Qy 381 LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 400
Db 3270 CTGAACAACGGAAGTCAAGCGGTGGGAGCGCTCATCTCTTTACTGCTCGAGTACTCCCT 3329
Qy 401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 420
Db 3330 TCGCAGATGTCTAGGACTGGAAATACTTCCAAATTCAGCTATACCTTTTCGAGGATGTACCT 3389
Qy 421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
Db 3390 TTTTCACAGCAGCTACGCTCACAGCAGAGTTGGATCGCTTGATGATCTCTTATTGAT 3449
Qy 441 GlnTyrLeuTyrTyrLeuAsnArgThrGln--AsnGlnSerGlySerAlaGlnAsnLys 459
Db 3450 CAGTATCTGTACTACTTGAACAGAACGCAAGGAACAACCTCTCGGAACAACCAACAATCA 3509
Qy 460 AspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeu 479
Db 3510 CGGCTGCTTTTTCAGCGGCTGGGCTCAGTCTATGCTTTTTCAGGCGCAGAAATTCGCTA 3569
Qy 480 ProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSer 499
Db 3570 CTTGGGCTCTTACCGGCAACAGAGACTTTCAAAGACTGCTAACGACAACAACAACAGT 3629
Qy 500 AsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsn 519
Db 3630 AACTTCTCTTGGACAGCGGCGCAAAATATCATCTCAATGGCCGCGACTCGCTGGTGAAT 3689
Qy 520 ProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGly 539
Db 3690 CCAGACACGATATGGCGACTCACAGACGATGAGAGAAATTTTTCCTATGCACGGC 3749
Qy 540 ValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMet 559
Db 3750 AATCTAATATTTGGCAAGGAGGCAACGCAAGTAACGCAAGATTAGATAATGTAATG 3809
Qy 560 IleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThr 579
Db 3810 ATTACGATGAAGAGAGATTCTGACCACCAATCTGTGGCAACGAGACGATATGGAAT 3869
Qy 580 ValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMet 599
Db 3870 GTGGCAATAACTTTCAGAGCTCAATATACACTCCACGACTGGACTGTCAATCATCAG 3929
Qy 600 GlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrp 619
Db 3930 GGGGCTTTTACCTGGCATGGTGTGGCAAGATCGTGACGTGTACTTCAAGGACCTATCTGG 3989
Qy 620 AlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGly 639
Db 3990 GCAAGATTCCTACAGGATGGACACTTTCATCTCTCTCTGTATGGGAGCTTTGGA 4049
Qy 640 LeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsnProPro 659
Db 4050 CTGAAACATCGCTCTCTCAATCATGATCAAAATACTCCCGTACCTCGCAATCTTCG 4109
Qy 660 AlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnVal 679
Db 4110 ACCACCTTCAGTGGGCAAGTTTGTCTTCTTCATCACACAGTACTCCACGGGACAGTTC 4169
Qy 680 SerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluVal 699
Db 4170 AGCTGGAGATCGAGTGGAGCTGCGAGGAGAAACAGCAACGCTGGAAATCCCGAAAT 4229
Qy 700 GlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGly 719
Db 4230 CAGTACACTTCCAACTACACAGTCTGTTAATCGTGGACTTACCGTGGATTAATGGC 4289
Qy 720 LeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 4290 GTGTATTTCAGAGCTTCGCCCTTCGCAATTTGGCACAGATACCTGACTCGTAATCTG 4340

RESULT 2

US-09-770-315-4
; Sequence 4, Application US/09770315
; Patent No. 6429001
; GENERAL INFORMATION:
; APPLICANT: Chiron Corporation
; TITLE OF INVENTION: Recombinant AAV Packaging Systems
; FILE REFERENCE: 20263-501
; CURRENT APPLICATION NUMBER: US/09/770,315
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,536
; PRIOR FILING DATE: 2000-01-26
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 4072
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: recombinant DNA
US-09-770-315-4

Alignment Scores:
Pred. No.: 0 Length: 4072
Score: 3402.50 Matches: 613
Percent Similarity: 90.22% Conservatives: 51
Best Local Similarity: 83.29% Mismatches: 71
Query Match: 85.30% Indels: 1
DB: 4 Gaps: 1

US-09-807-802A-13 (1-736) x US-09-770-315-4 (1-4072)

QY 1 MetAlaAlaAspGlyTyrLeuProAspTrpLeuGluAspAsnLeuSerGluGlyIleArg 20
DB 1484 ATGGCTGCGATGGTATCTTCCAGATTGGCTCGAGGACACTCTCTCTGAAGGATAGA 1543
QY 21 GluTrpTrpAspLeuProGlyAlaProLysProLysAlaAsnGlnGlnLysGlnAsp 40
DB 1544 CAGTGGTGGAGCTCAACCTGGCCACCACCACCACCAAGCCGCGAGCGGATAGGAC 1603
QY 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
DB 1604 GACAGCGGGGCTTGTGCTCTCTGGGTACAGTACCTCGGACCTTCAACGGACTCGAC 1663
QY 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
DB 1664 AAGGAGAGCCGGTCAACGAGCAGACGCCGCGGCCCTCGAGCAGCAAGAGCTACGAC 1723
QY 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
DB 1724 CGGCAGCTCGACAGCGGAGACAACCCGTACCTCAAGTACAACACCAGCCGCGAGGTT 1783
QY 101 GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln 120
DB 1784 CAGGAGCGGCTTAAAGAGATACGCTTTTGGGGGCAACCTCGGACGAGCAGTCTTCAG 1843
QY 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaPro 140
DB 1844 GCGAAAGAGGGTCTTGAACCTCTGGGCTGGTTGAGGACCTGTTAAGACGGCTCG 1903
QY 141 GlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerGlyIleGly 160
DB 1904 GGAAGAGAGCGCGGTAGAGCACTCTCTGTGGAGCCAGACTCTCTCTCGGGAACCGGA 1963
QY 161 LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
DB 1964 AAGCGGGCGCCAGCGCTTCAAGAAAAAGATTGAATTTTGGTCAGCAGTGGAGACCGAC 2023
QY 181 SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro 200
DB 2024 TCAGTACCTGACCCCGAGCTCTCGGACAGCCACCAGCAGGCCCTCTGGTCTGGGAAC 2083

QY 201 ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnGluGlyAlaAspGly 220
DB 2084 AATACGATGGCTACAGGCGAGTGGCCACCAATGGCAGACAATAACGAGGGCGCGACGGA 2143
QY 221 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIle 240
DB 2144 GTGGGTAAATTCCTCGGAAATTTGGCATTTGCCATTTCCATGGATGGGCGCAGAGTCAATC 2203
QY 241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIle 260
DB 2204 ACCACGAGCACCCGNACTGGGCCCTGCCACCTACCAACAACCACTCTACAAACAAT 2263
QY 261 SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrp 280
DB 2264 TCCAGCCCAATCA--GGAGCTCGAAGCAGCAATCACTACTTTGGCTACAGCACCCCTGG 2320
QY 281 GlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu 300
DB 2321 GGGTAATTTTGCATTTCAACAGATTCCACTGCGACTTTTCCACACGCTGACTGGCAAGACTC 2380
QY 301 IleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGln 320
DB 2381 ATCAACAACAACCTGGGATTTCCGACCCCAAGAGACTCAACTTCAAGCTCTTTAACTCAA 2440
QY 321 ValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThr 340
DB 2441 GTCAAGAGGTCACGCAAGATGACGGTACGACGACGATGGCCATTAACCTTTACAGCAGC 2500
QY 341 ValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGln 360
DB 2501 GTTCAGGTGTTTACTGACTCGAGTACAGCTCCGCTCCGCTCGGCTCGGCGCAGTCA 2560
QY 361 GlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr 380
DB 2561 GGATGCTCCCGCGGTTCCCGACGACAGCGTCTTCATGTGGCCACAGTAGTAGTACTCACC 2620
QY 381 LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 400
DB 2621 CTGAACAACGGGAGTCAGGAGTAGGAGCGCTCTTCTATTTACTCGCTGGAGTACTTCTCT 2680
QY 401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 420
DB 2681 TCTCAGATGCTGCGTACCGGAAACAATTTTACCTTCAGCTACACTTTTGGAGCGTTCTCT 2740
QY 421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
DB 2741 TTCCACAGCAGCTACGCTCACAGCAGAGTCTGACCGCTCTCATGAATCTCTCATCGAC 2800
QY 441 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAsp 460
DB 2801 CAGTACCTGTATTACTTTGAGCAGAAACAACACTCCAAGTGGAAACCCACCGACGCTCAAG 2860
QY 461 LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuPro 480
DB 2861 CTTCAGTTTCTCAGCGCGGAGCGAGTGCACATTGGGACCCAGCTTAGGAATCGCTTCTCT 2920
QY 481 GlyProCysTyrArgGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsn 500
DB 2921 GGACCTCTTACCAGCAGCAGAGTATCAAGACATCTCGCGGATTAACAACAACAGAGTAA 2980
QY 501 PheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnPro 520
DB 2981 TACTCGTGAGCTGAGCTTACCAGTACCTCAATGGCAGAGACTCTCTGGTGAATCCG 3040
QY 521 GlyThrAlaMetAlaSerHisLysAspGluAspLysPhePheProMetSerGlyVal 540
DB 3041 GGCCCGCCCATGGCAGCCACAGCAGATGAAGAAAGTTTCTCTCAGAGCGGGT 3100
QY 541 MetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIle 560
DB 3101 CTCATCTTTGGAGAGCAGGCTCAGAGAAAAAATAATGTGGACATTGAAAAGGTCTATGATT 3160
QY 561 ThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrVal 580


```
Qy 401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 420
Db 3428 TCTCAGATGCTGGGTACCGAACAACACTTTACCTTCAGTACACTTTTGAGAGAGTTCCT 3487
Qy 421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
Db 3488 TTCCACAGCAGCTACGCTCACAGCCAGAGTCTGGACCGTCTCATGAATCTCTCATCGAC 3547
Qy 441 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAsp 460
Db 3548 CAGTACCTGTATTACTTACGAGAGAACAAACACTCCAAAGTGAACCCACCGAGTCAAGG 3607
Qy 461 LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuPro 480
Db 3608 CTTCACTTTTCTCAGCCCGGAGTGCATTCGGGACAGTCTAGGAACCTGGCTTCCT 3667
Qy 481 GlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsn 500
Db 3668 GGACCTCTGTACCGCCAGCAGCGAGTATCAAGACATCTGGCGATTAACAACAACAGTGAA 3727
Qy 501 PheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleLeuAsnPro 520
Db 3728 TACTCGTGAGTGGAGTACCAAGTACCACCTCAATGGCAGAGACTCTCTGGTGAATCCG 3787
Qy 521 GlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyVal 540
Db 3788 GGCCCGGCCATGGCAAGCCACAGACGATGAGAAAGTTTTCTCTCAGAGCGGGTT 3847
Qy 541 MetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIle 560
Db 3848 CTCATCTTTGGAGAGAGGCTTCAGAGAAACAAATGTGGACATTTGAAAGGTTCATGATT 3907
Qy 561 ThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrVal 580
Db 3908 ACAGACGAGAGAGAAATCAGGACAAACCAATCCCGTGGCTACGAGCAGATGTTCTGTA 3967
Qy 581 AlaValAsnPheGlnSerSerSerProAlaThrGlyAspValHisAlaMetGly 600
Db 3968 TCTACCAACCTCCAGAGAGCCACAGACAGCAGCTACCGCAGATGTCAACACACAAGGC 4027
Qy 601 AlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAla 620
Db 4028 GTTCTTCCAGGCATGCTCTGGCAGACACAGATGTGTACCTTCAGGGGCCCATCTGGCA 4087
Qy 621 LysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeu 640
Db 4088 AAGATTCCACACACGAGCGACATTTTACCCTCTCCCTCATGGGTGGATTCGAGCTT 4147
Qy 641 LysAsnProProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAla 660
Db 4148 NAACACCTCTCTCCACAGATTCTCATCAAGAACACCCCGTACCTGCGAATCTTTCGACC 4207
Qy 661 GluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSer 680
Db 4208 ACCTTCAGTCGCGCAAGTTTCTCTTCATCACACAGTATCTCCACGGGACAGGTGACG 4267
Qy 681 ValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTyrAsnProGluValGln 700
Db 4268 GTGGAGATCGAGTGGAGCTGAGAGAGGAAACAGCAACGCTGGAATCCCGAATTCAG 4327
Qy 701 TyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeu 720
Db 4328 TACACTTCAACTACAAACAGTCTGTTAATGTGGACTTTACTGTGGACACTAATGCGGTG 4387
Qy 721 TyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 4388 TATTACAGACCTTCGCCCCATTGGCCACCATGATCTGACTGCTGAATCTG 4435
```

RESULT 4

US-09-770-315-2

; Sequence 2, Application US/09770315

; Patent No. 6429001

; GENERAL INFORMATION:

; APPLICANT: Chiron Corporation
; TITLE OF INVENTION: Recombinant AAV Packaging Systems
; FILE REFERENCE: 20263-501

; CURRENT APPLICATION NUMBER: US/09/770,315

; PRIOR FILING DATE: 2001-01-26

; PRIOR APPLICATION NUMBER: US 60/178,536

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 8698

; TYPE: DNA

; ORGANISM: Unknown

; FEATURE:

; OTHER INFORMATION: recombinant DNA

US-09-770-315-2

Alignment Scores:

| | | | |
|------------------------|---------|---------------|------|
| Pred. No.: | 0 | Length: | 8698 |
| Score: | 3402.50 | Matches: | 613 |
| Percent Similarity: | 90.22% | Conservative: | 51 |
| Best Local Similarity: | 83.29% | Mismatches: | 71 |
| Query Match: | 85.30% | Indels: | 1 |
| DB: | 4 | Gaps: | 1 |

US-09-807-802A-13 (1-736) x US-09-770-315-2 (1-8698)

```
Qy 1 MetAlaAlaAspGlyTyrLeuProAspTrpLeuGluAsnLeuSerGluGlyLeuArg 20
Db 2203 ATGCTCTCCCATGTTTATCTTCAGATGTGGCTCAGGACACTCTCTCTGAAGGAATAGA 2262
Qy 21 GluTrpTrpAspLeuLysProGlyAlaProLysProLysAlaAsnGlnLysGlnAsp 40
Db 2263 CAGTGGTGGAGCTCAAACTGGCCACCAACCAAGCCCGCAGAGCGCGAATAAGGAC 2322
Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
Db 2323 GACAGCAGGGGTCTTGTGCTTCTGGGTACAAAGTACCTCGGACCTTCAACGGACTCGAC 2382
Qy 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
Db 2383 AAGGAGAGCGCGGTCAACAGGCGAGACGCCCGGCCCTTCGAGCAGCAACAAAGCTTACGC 2442
Qy 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
Db 2443 CGGCAGCTCGACAGCGGAGACACCCGTACCTCAAGTACACCAACCGCGCGAGGTTT 2502
Qy 101 GlnGluArgLeuGlnLysAspThrSerPheGlyAsnLeuGlyArgAlaValPheGln 120
Db 2503 CAGGAGCGCTTAAAGAAGATACGTCTTTTGGGGCAACCTTCGGACGAGCAGTCTTCCAG 2562
Qy 121 AlaLysLysArgValLeuGluProLeuGluValGluGluGlyAlaLysThrAlaPro 140
Db 2563 GCGAAAAGAGGGTCTTGAACCTCTGGGCTGGTTGAGGAACTGTTAAGACGGCTCCG 2622
Qy 141 GlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSerGlyIleGly 160
Db 2623 GGAANAAGAGCGCGGTAGAGCACTCTCTCTGTGGAGCCAGACTCTCTCTCGGAGCGAGC 2682
Qy 161 LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
Db 2683 AAGCGGGCCAGCAGCTGCAAGAAAAGATTGAATTTTGGTTCAGACTGGAGAGCGCAGAC 2742
Qy 181 SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro 200
Db 2743 TCAGTACTGACCCCGAGCTCTCGGACAGCCACAGCAGCCCTCTGCTCTGGGAAC 2802
Qy 201 ThrThrMetAlaSerGlyGlyValaProMetAlaAspAsnAsnGluGlyAlaAspGly 220
Db 2803 AATACGATGGCTACAGGACGTGGCGCACCACCAATGCGCAGACATAACGAGGGCGCGAC 2862
Qy 221 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIle 240
```

Db 2863 GTGGGTAATTCCTCGGGAATTTGGCAATTCGATTCACATGATGGGACAGAGTCATC 2922
Qy 241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrIleGlnle 260
Db 2923 ACCACGACACCGAACCTGGGCGCTCCGACCTACCAACCAACCACTCTACAAACAAAT 2982
Qy 261 SerSerAlaSerThrGlyAlaSerAsnAsnHisTyrPheGlyTyrSerThrProTrp 280
Db 2983 TCCAGCCAAATCA--GGAGCCCTCGAACGACAAATCACTACTTTGGCTACAGCACCCCTTGG 3039
Qy 281 GlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu 300
Db 3040 GGGTATTTTGACTTCAACAGATTCCTCCGACCTTTCCACGAGTGAGTGGCAAGAGCTC 3099
Qy 301 IleAsnAsnAsnTrpGlyPheArgProTyrArgLeuAsnPheLeuPheAsnIleGln 320
Db 3100 ATCAACAACTGGGGATTCGACCCCAAGAGACTCAACTTCAAGCTCTTTAAACATTCAA 3159
Qy 321 ValIleGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThr 340
Db 3160 GTCAAGAGAGGTCAAGAGATGACGGTACGACGAGATTCGCAATTAACCTTACCAAGCAG 3219
Qy 341 ValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGln 360
Db 3220 GTTCAGGTGTTTACTGACTCGGAGTACGAGCTCCGCTACGTCTCGGCTCGGCGCATCA 3279
Qy 361 GlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr 380
Db 3280 GGATGCTCCCGCGGTTCCAGCAGACGCTCTTCATGGTGCCACAGTATGGATACCTCACC 3339
Qy 381 LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 400
Db 3340 CTGAACACCGGAGTCAGGAGTAGGAGCGCTCTTCATTTTACTGCTCGGAGTACTTCTCT 3399
Qy 401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 420
Db 3400 TCTCAGATGCTGGTACCGGAAACACTTTTACCTTCAGCTACACTTTTCGAGGAGCTTCT 3459
Qy 421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProIleuLeuAsp 440
Db 3460 TTCCACAGCAGCTACGCTCACAGCCAGAGTCTGGACCGCTCTCATGAATCTCTCATCGAC 3519
Qy 441 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnIleAsp 460
Db 3520 CAGTACTGTATTACTTGAGCAGACAAACACTCCAGTGGACACCCACGACGAGTCAAGG 3579
Qy 461 LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProIleAsnTrpLeuPro 480
Db 3580 CTTTCAGTTTCTCAGCGCGGAGGAGTGACATTCGGGACAGTCTAGGAACCTGGCTTCT 3639
Qy 481 GlyProCysTyrArgGlnGlnArgValSerIleThrLysThrAspAsnAsnAsnSerAsn 500
Db 3640 GGACCCCTGTTCACCGCCAGCAGCGAGTATCAAGACATCTCGCGGATCAACCAACAGTGA 3699
Qy 501 PheThrTrpThrGlyAlaSerIleTyrAsnLeuAsnGlyArgGluSerIleIleAsnPro 520
Db 3700 TACTGTGGACTGGAGCTACCAAGTACCTCCATGGAGAGACTCTCTGGTGAATCCG 3759
Qy 521 GlyThrAlaMetAlaSerHisIleAspAspGluAspPhePheProMetSerGlyVal 540
Db 3760 GGCGCCGCGCATGCCAAGCCACAGGACGATGAAGAAAGTTTCTTCAGAGCGGGT 3819
Qy 541 MetIlePheGlyIleGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIle 560
Db 3820 CTCATCTTTGGGAAGCAAGCTCAGAGAAACAAATGTGGACATGAAAAGGTCTAGT 3879
Qy 561 ThrAspGluGluIleLeuAlaThrAsnProValAlaThrGluArgPheGlyThrVal 580
Db 3880 ACAGACGAGAGGAATCAGGACAAACATCCGCTGCTACGGAGCAGTATGTTCTGTA 3939
Qy 581 AlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGly 600

Db 3940 TCTACCACTCCAGAGGCGCAACAGCAGCTACCGAGATGTCAACACACACAAGGC 3999
Qy 601 AlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAla 620
Db 4000 GTTCTTCAGGATGGTCTGGCAGCAGAGATGTACCTTCAGGGGCCCATCTGGGCA 4059
Qy 621 LysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeu 640
Db 4060 AAGATTTCACACACGAGCGGACATTTTCCACCCCTCTCCCTCATGGGTGGATTGGACTT 4119
Qy 641 LysAsnProProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAla 660
Db 4120 AAACACCTCTCCACAGATTTCTCATCAAGAACACCCCGGTACCTGGGAATCTTCGACC 4179
Qy 661 GluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGluValSer 680
Db 4180 ACCTTCAGTGGCGCAAGTTTGTCTTCATCATCAGATCTCCACGGGACAGGTACGC 4239
Qy 681 ValGluIleGluTyrGluLeuGlnLysGluAsnSerLysArgTyrAsnProGluValGln 700
Db 4240 GTGGAGATCGAGTGGAGCTGCAGAGGAAACAGCAACCGCTGGAAATCCCGAAATTCAG 4299
Qy 701 TyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnGlyLeu 720
Db 4300 TACACTTCCAACTACAAACAGTCTGTTAATGTGACCTTTACTGTGGACACTAATGGCGTG 4359
Qy 721 TyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 4360 TATTCAGAGCTCGCCCATTTGGCACCAGATACCTGACTCGTAATCTG 4407

RESULT 5

US-08-254-358-1
; Sequence 1, Application US/08254358
; Patent No. 5658785
; GENERAL INFORMATION:
; APPLICANT: Johnson, Philip R.
; TITLE OF INVENTION: Adeno-Associated Virus Materials and
; TITLE OF INVENTION: Methods
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/254,358
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5658785and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31975
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4680 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-254-358-1

Alignment Scores:

| | | | |
|------------------------|-----------|---------------|------|
| Pred. No.: | 9.71e-315 | Length: | 4680 |
| Score: | 3360.50 | Matches: | 609 |
| Percent Similarity: | 89.42% | Conservative: | 50 |
| Best Local Similarity: | 82.63% | Mismatches: | 76 |
| Query Match: | 84.24% | Indels: | 2 |
| DB: | 1 | Gaps: | 1 |

US-09-807-802A-13 (1-736) x US-08-254-358-1 (1-4680)

| | | | |
|----|------|---|------|
| Qy | 1 | MetAlaAlaAspGlyTyrLeuProAspTrpLeuGluAspAsnLeuSerGluGlyIleArg | 20 |
| Db | 2203 | ATGGCTGGCGATGGTATCTTCCAGATTGGCTCGAGGACACTCTCTCTGAGGAATAAGA | 2262 |
| Qy | 21 | GlutTrpAspLeuLeuProGlyAlaProLysProLysAlaAlaAsnGlnGlnLysGlnAsp | 40 |
| Db | 2263 | CAGTGGTGGAGCTCAACCTGGCCACCACCACCACCAAGCCGCGAGCGGCAATAGGAC | 2322 |
| Qy | 41 | AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp | 60 |
| Db | 2323 | GACAGCGGGGCTTGTGCTTCTGGGTACAAATGACTCTCGGACCTTCAACGGACTCGAC | 2382 |
| Qy | 61 | LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp | 80 |
| Db | 2383 | AAGGAGAGCCGCTCAACGAGCGACAGCCGCGGCGCTCGAGCACCAAGGCTTACGAC | 2442 |
| Qy | 81 | GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe | 100 |
| Db | 2443 | CGCGAGCTCGACAGCGGAGACAACCGTACTCTCAAGTACAACCAACCGCGCGAGTTT | 2502 |
| Qy | 101 | GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln | 120 |
| Db | 2503 | CAGGAGCGCTTAAAGAGATACGTCTTTTGGGGGCAACCTCGGACGACGAGTCTTCCAG | 2562 |
| Qy | 121 | AlaLysLysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaPro | 140 |
| Db | 2563 | GGCAAAAGAGGGTCTTGAACTCTGGCCCTGGTTGAGGAACCTGTTAAGACGGCTCG | 2622 |
| Qy | 141 | GlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerGlyIleGly | 160 |
| Db | 2623 | GGAAAAAGAGCGGTGAGCACCTCTCTGTGGAGCCAGACTCTCTCTCGGGAACCGGA | 2682 |
| Qy | 161 | LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu | 180 |
| Db | 2683 | AAGCGGGGCGACAGCCTTCAGACCAAAAGATTGAAATTTTGGTCAGACTGGAGACCGAC | 2742 |
| Qy | 181 | SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro | 200 |
| Db | 2743 | TCAGTACCTTGACCCCGAGCTCTCGGACAGCCACCAGCCCTCTCTGGTCTGGGAAT | 2802 |
| Qy | 201 | ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnGluGlyAlaAspGly | 220 |
| Db | 2803 | AATACGATGGCTACAGCGAGTGGCGCACCAATGGCAGACAAATAACGAGGGCGCGACGGA | 2862 |
| Qy | 221 | ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIle | 240 |
| Db | 2863 | GTGGGTAATTTCTCCGGAAATTTGGCATTCGGATTCCCATGATGGATGGCGACAGTCAATC | 2922 |
| Qy | 241 | ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLysLeuTyrLysGlnIle | 260 |
| Db | 2923 | ACCACGACCCCGAACTTGGGCCCTGGCCCACTACACCAACCACTCTACAAACAAT | 2982 |
| Qy | 261 | SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrp | 280 |
| Db | 2983 | TCCAGCCAATCA--GGAGCTCGAACGACCAATCACTACTTTGGCTACAGCACCCCTTGG | 3039 |
| Qy | 281 | GlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu | 300 |
| Db | 3040 | GGGTATTTTGACTTCAACAGATTCCACTGCTCCACTTTTCCACCAAGTGGCAAAAGACTC | 3099 |
| Qy | 301 | IleAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGln | 320 |
| Db | 3100 | ATCAACACACACTGGGGATTCGACCCCAAGAGACTCACTTCAAGCTCTTTAACAATCA | 3159 |

| | | | |
|----|------|---|------|
| Qy | 321 | ValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnLeuThrSerThr | 340 |
| Db | 3160 | GTCAAGAGGTCACCGCAGATGACGGTACGACACGAGTATGCCATAAATCTTACGACACG | 3219 |
| Qy | 341 | ValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGln | 360 |
| Db | 3220 | GTTCAGGTGTTTACTGACTCGGAGTACAGTCCCGTACGTCCTCGGCTCGGCGATCAAA | 3279 |
| Qy | 361 | GlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr | 380 |
| Db | 3280 | GGATGCTCCCGCGTTCCTCCGACGACAGCTCTTCATGGTGCCACAGTATGGATACCTCAC | 3339 |
| Qy | 381 | LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro | 400 |
| Db | 3340 | CTGAACAACGGGAGTCAGGAGTAGGACGCTCTTTCATTTACTCGCTCGGAGTACTTCCT | 3399 |
| Qy | 401 | SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro | 420 |
| Db | 3400 | TCTCAGATGCTGCGTACCGGAACACTTTTACCTTCAGCTACACTTTTGAGGAGCTTCCT | 3459 |
| Qy | 421 | PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp | 440 |
| Db | 3460 | TTCCACAGCAGCTACGCTCACAGCCAGAGTCTGGACCGTCTCATGAATCTCTCATCGAC | 3519 |
| Qy | 441 | GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAsp | 460 |
| Db | 3520 | CAGTACCTGTATTTACTTTGAGCAGAACCAACTCCAAGTGGAAACCCACACGAGTCAAGG | 3579 |
| Qy | 461 | LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuPro | 480 |
| Db | 3580 | CTTCAGTTTTCTCAGCGCGGAGCGGTGACATTCGGGACCGAGTCTAGGAACTGGCTTCCT | 3639 |
| Qy | 481 | GlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsn | 500 |
| Db | 3640 | GGACCTGTTTACCGCCAGCAGCGAGTATCAAGACATCTCGCGGATAACAACAACAGTGAA | 3699 |
| Qy | 501 | PheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnPro | 520 |
| Db | 3700 | TACTCGTGAGCTGAGACTACCAAGTACACCTCAATGGCAGAGACTCTCTCGTGAATCCG | 3759 |
| Qy | 521 | GlyThrAlaMetAlaSerHisLysAspGluAspLysPhePheProMetSerGlyVal | 540 |
| Db | 3760 | GGGCGCCGATGGACGCCACAGACGATGAGAAAGATTTTTTCTCAGAGCGGGGTT | 3819 |
| Qy | 541 | MetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIle | 560 |
| Db | 3820 | CTCATCTTTGGAGCAAGGCTCAGAGAAAACAAATGTGAACATTTGAAAAGGTTCATGAT | 3879 |
| Qy | 561 | ThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrVal | 580 |
| Db | 3880 | ACAGACGAAGAGGAAATCGGAACCAACCAATCCCGTGGCTACGGAGCATGTGGTTCTGTA | 3939 |
| Qy | 591 | AlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGly | 600 |
| Db | 3940 | TCTACCAACTCCAGAGGGCAACAGACAGCTACCGCAGATGTCAACACACAAAGGC | 3999 |
| Qy | 601 | AlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAla | 620 |
| Db | 4000 | GTTCTTCAGCATGCTCTGGCAGACAGAGATGTGTACCTTCAGGGGCCCATCTGGGCA | 4059 |
| Qy | 621 | LysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyPheGlyLeu | 640 |
| Db | 4060 | AAGATTCCACACAGGACGGACATTTTCCACCCCTCTCCCTCATGGGTGGATTCGACTT | 4119 |
| Qy | 641 | LysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAla | 660 |
| Db | 4120 | AAACACCTCTCCACAGATTTCTCATCAAGAAACACCCCGGTACCTCGGAATCTTCGACC | 4179 |
| Qy | 661 | GluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGln-ValIse | 680 |
| Db | 4180 | ACCTTCAGTCGGCAAGTTTGTCTCTTCATCACAGTAGTCTCCACGGGACACGCTCAG | 4239 |
| Qy | 680 | rValGluIleGluTrpGluLeuGlnLysSerLysArgTrpAsnProGluValGlu | 700 |

Db 4240 CTGGAGATCGAGTGGGAGCTCGAGAGGAAACAGAAACCTCGAATCCCGAATTC 4299
Qy 700 nTyrThrSerAntyAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLe 720
Db 4300 GTACACTTCCAACTCAACAACTCTGTTAATCGTGGACTTACCGTGGTACTAATGGCGT 4359
Qy 720 uTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 4360 GTATTTCAGAGCTCGCCCAATGGCACCAACATACCTGACTCGTAATCTG 4408

RESULT 6

US-08-475-391-1
Sequence 1, Application US/08475391
Patent No. 5786211
GENERAL INFORMATION:
APPLICANT: Johnson, Philip R.
TITLE OF INVENTION: Adeno-Associated Virus Materials and
METHODS
TITLE OF INVENTION: Methods
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSES: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 S. Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,391
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/254,358
FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: No. 578621land, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31975
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 4680 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-475-391-1

Alignment Scores:

Pred. No.: 9,716-315 Length: 4680
Score: 3360.50 Matches: 609
Percent Similarity: 89.42% Conservativeness: 50
Best Local Similarity: 82.63% Mismatches: 76
Query Match: 84.24% Indels: 2
DB: 1 Gaps: 1

US-09-807-802a-13 (1-736) x US-08-475-391-1 (1-4680)

Qy 1 MetAlaAlaAspGlyTyrLeuProArgTrpLeuGluAspAsnLeuSerGluGlyLeArg 20
Db 2203 ATGGCTGCCGATGGTTATCTCTCCAGATTGGCTCGAGGACACTCTCTCTGAAGGAATAAGA 2262
Qy 21 GluTrpTrpAspLeuIysProGlyAlaProLysProIysAlaAsnGlnGlnIysGlnAsp 40
Db 2263 CAGTGGTGGGAAGCTCAAAACCTGGCCCAACCAACCAAGCCGACGAGCGGCATTAAGGAC 2322

Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrIysTyrLeuGlyProPheAsnGlyLeuAsp 60
Db 2223 GACAGCAGGGGTCTTGTGCTTCTGGGTACAGTACCTCGGACCTTCAACGGACTCGAC 2382
Qy 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaIysAsp 80
Db 2383 AAGGGAGGCGGTCAACAGGAGGAGCGCGCGCCCTCGAGCAGCAGCAAGCCTACGAC 2442
Qy 81 GlnGlnLeuIysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
Db 2443 CGGAGCTCGACAGCGGAGACAACCGTACCTCAAGTACACACCGCGCGCGGAGTTT 2502
Qy 101 GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyAspAlaValPheGln 120
Db 2503 CAGGAGCGCTTAAAGAAGATACGTCTTTGGGGGCAACCTCGGACGAGCAGTCTTCAG 2562
Qy 121 AlalysIysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaPro 140
Db 2563 GCGAANAAGAGGGTCTTGAACCTCTGGGCTGTTGAGGAACCTGTTAAGACGGCTCG 2622
Qy 141 GlyIysIysArgProValGluGlnSerProGlnGluProAspSerSerSerGlyIleGly 160
Db 2623 GGAANAAGAGCGGTAGAGCACTCTCTGTGGAGCCAGACTCTCTCTCGGGAACCGGA 2682
Qy 161 LysThrGlyGlnGlnProAlaLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
Db 2683 AAGGGGCGCCAGCAGCTCTGCAAGAAAGATTGAATTTGGTCACTCGGAGAGCGCAGAC 2742
Qy 181 SerValProAspProGlnProLeuGlyGluProAlaThrProAlaAlaValGlyPro 200
Db 2743 TCAGTACCTGACCCCGGAGCTCTCGGACGACCAACAGCGAGCCCTCTCTGTTGGAAT 2802
Qy 201 ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGly 220
Db 2803 AATACGATGGCTACAGGAGTGGCGCACCAATGGCAGACAATACAGGGCGCGCCGGA 2862
Qy 221 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIle 240
Db 2863 GTGGGTAAATCTCTCGGNAATTTGGCATTTGCATTTCCACATGATGGGCGCAGAGTCATC 2922
Qy 241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrIysGlnIle 260
Db 2923 ACCACGACGCCGAAACCTGGGCGCTGCCACCTACACCAACCAACCTCTACAAACAAAT 2982
Qy 261 SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrp 280
Db 2983 TCCAGCCCAATCA---GGAGCCTCGAACCGACAATCACTACTTTGGCTACGACACCCCTTGG 3039
Qy 281 GlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu 300
Db 3040 GGGTATTTTGTACTTCAACAGATTCCACTGCCACTTTTCCACCGGTGACTGGCAAGACTC 3099
Qy 301 IleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGln 320
Db 3100 ATCAACAACAATCGGGATTCCGACCCCAAGAGACTCAACTTCAAGCTCTTTAACAATTCAA 3159
Qy 321 ValIysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThr 340
Db 3160 GTCAAAAGAGGTCAACAGAAATGACGGTACGACGAGCATTTGCCAATTAACCTTACCGAC 3219
Qy 341 ValGlnValPheSerAspSerGlyTyrGlnLeuProTyrValLeuGlySerAlaHisGln 360
Db 3220 GTTCAGGTGTTTACTGACTCGGAGTACCGACTCCCGTACGCTCTCGGCTCGGCGCATCAA 3279
Qy 361 GlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr 380
Db 3280 GGATGCCTCCCGCGCTTCCAGCAGACGTCTTCATGGTGGCCACAGATATGGATACCTCAC 3339
Qy 381 LeuAsnAsnGlySerGlnAlaValGlyArgSerPheTyrCysLeuGluTyrPhePro 400
Db 3340 CTGAACAACGGAGTCAAGCAGTAGAGCGCTCTTCAATTTTACTGCTGGAGTACTTCTCT 3399

```
Qy 401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyThrPheGluGluValPro 420
Db 3400 TCTCAGATGCTGCTGACCGAAACAACTTTACCTTACGTACACTTTTGGAGCGTTTCT 3459
Qy 421 PheHisSerSerTyAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuLeuAsp 440
Db 3460 TTCCACAGCAGCTACGCTCACAGCGAGAGTCTGGACCGTCTCATGAATCTCTCATCGAC 3519
Qy 441 GlnTyLeuTyTyLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAsp 460
Db 3520 CAGTACCTGTATTACTTGAGCAGAGAACAACTCCNAGTGAACACCCAGCAGCTCAAGG 3579
Qy 461 LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuPro 480
Db 3580 CTTTCAGTTTCTCAGGCGGAGCGAGTACATTCGGGACCACTAGTACGAACTGCTTCTCT 3639
Qy 481 GlyProCysTyArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsn 500
Db 3640 GGACCTCTGTACCCGCGAGCAGAGTATCAAGAGATCTGCGGATAACAACACAGTGAA 3699
Qy 501 PheTrpTrpThrGlyAlaSerLysTyAsnLeuAsnGlyArgGluSerIleIleAsnPro 520
Db 3700 TACTCGTGAGCTGAGCTACCAAGTACCACTCAATGGCAGAGACTCTCTGGTGAATCCG 3759
Qy 521 GlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyVal 540
Db 3760 GGGCCCGCATGCGCAAGCCACAGGACGATGAAGAAAGTTTTCCTTCAGAGCGGGTT 3819
Qy 541 MetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIle 560
Db 3820 CTCATCTTTGGAGCAGGCTCAGAGAAACAAATGTGAACATTTGAAGAGTTCATGATT 3879
Qy 561 ThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrVal 580
Db 3880 ACAGACGAAGAGGAATCGGAACCAACCAATCCCGTGGCTACGAGCAGTATGTTCTGTA 3939
Qy 581 AlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGly 600
Db 3940 TCTACCACTCCAGAGGCAACAGACAAGCAGCTACCGCAGATGTCAACACACAAGGC 3999
Qy 601 AlaLeuProGlyMetValTrpGlnAspArgAspValTyLeuGlnGlyProIleTrpAla 620
Db 4000 GTTCTTCAGGATGCTGTCGAGCAGACAGATGTACCTTCAGGGGCCCATCTGGGCA 4059
Qy 621 LysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeu 640
Db 4060 AAGATTCCACACCGGACGACATTTTCAACCTCTCCCTCATGGTGGATTCGACTT 4119
Qy 641 LysAsnProProProGlnIleLeuLysAsnThrProValProAlaAsnProProAla 660
Db 4120 AAACACCTCTCCACAGATTTCTATCAAGAACACCCCGGTACCTGCGAATCTCTCGACC 4179
Qy 661 GluPheSerAlaThrLysPheAlaSerPheIleThrGlnTySerThrGlyGln-ValSe 680
Db 4180 ACCTTCAGTGGCGMAAGTTTGTCTTTCATCACAGTACTCCACGGGACCGGTACG 4239
Qy 680 rValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValG1 700
Db 4240 CGTGAGATCAGTGGGAGCTGCAGAGGAANAACAGCAACACGCTGGAATCCCGAAATCA 4299
Qy 700 nTyThrSerAsnTyAlaLysSerAlaAsnValAspPheThrValAsnAsnGlyLe 720
Db 4300 GTACACTTCCAACTACAAAGTCTGTTAATCTGGGACTTACCGTGGATATAATAATGGCGT 4359
Qy 720 uTyThrGluProArgProIleGlyThrArgTyLeuThrArgProLeu 736
Db 4360 GTATTACAGACTTCGCCCATTTGGACCAAGATACCTGACTGTAATCTG 4408
```

RESULT 7

US-08-709-609-1

; Sequence 1, Application US/08709609

; Patent No. 5858775

; GENERAL INFORMATION:

```
; APPLICANT: Johnson, Philip R.
; TITLE OF INVENTION: Adeno-Associated Virus Materials and
; METHOD OF INVENTION: Methods
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/709,609
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5858775and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31975
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4680 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-709-609-1

Alignment Scores:
Pred. No.: 9,71e-315 Length: 4680
Score: 3360.50 Matches: 609
Percent Similarity: 89.42% Conservative: 50
Best Local Similarity: 82.63% Mismatches: 76
Query Match: 84.24% Indels: 2
DB: Gaps: 1

US-09-807-802A-13 (1-736) x US-08-709-609-1 (1-4680)

Qy 21 MetAlaAspGlyTyLeuProAspTrpLeuGluAsnLeuSerGluGlyLeArg 20
Db 2203 ATGGCTGCCGATGGTTATCTTCAGATTGGCTCGAGGACACTCTCTCAAGGANTAGA 2262
Qy 21 GluTrpTrpAspLeuLysProGlyAlaProLysProLysAlaAsnGlnLysGlnAsp 40
Db 2263 CAGTGGTGGAGCTCAAACTGGCCACCACCAAGCCCGCAGAGCGCATAGGAC 2322
Qy 41 AspGlyArgGlyLeuValLeuProGlyTyTyLeuGlyProPheAsnGlyLeuAsp 60
Db 2323 GACAGCAGGGGTCTTGCTCTCTCGGTACAGTACCTCGGACCTTCAACGGGACTCGAC 2382
Qy 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyAsp 80
Db 2383 AAGGAGAGCGGTCAACAGGAGGAGCGCGCGGCTTCGAGCACGACCAAGCCCTACGAC 2442
Qy 81 GlnGlnLeuLysAlaGlyAspAsnProTyLeuArgTyAsnHisAlaAspAlaGluPhe 100
Db 2443 CGGCGAGCTCGACAGCGGAGACACCCGCTACTCAAGTACAAACACCGCGCGAGTTT 2502
Qy 101 GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln 120
Db 2503 CAGGAGCGCTTAAAGAAGATAGCTCTTTTGGGGGCAACCTCGGACGAGCAGTCTTCAG 2562
Qy 121 AlaLysLysArgValLeuGluProLeuGlyValGluGlyAlaLysThrAlaPro 140
```

Db 2563 GCGAAAAAGAGGTTCTTGAACCTCTGGGCTGGTGGAGAACCTGTTAAAGACGGCTCCG 2622
Qy 141 GlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSerGlyIleGly 160
Db 2623 GCGAAAAAGAGGCGGTAGACACTCTCTCTGGAGAGCCAGACTCTCTCTCGGAACCGGA 2682
Qy 161 LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
Db 2683 AAGGGCGGCCACGACGCTCTGCAAGAAAAGATTGAATTTTGGTTCAGACTGGAGACCGAC 2742
Qy 181 SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro 200
Db 2743 TCAGTACCTGACCCCGAGCTCTCGACAGCCACCGACGCCCCCTCTGGTCTGGGAAT 2802
Qy 201 ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnGluGlyAlaAspGly 220
Db 2803 AATACGATGGCTACAGGAGTGGCGCACCAATGGCGAGCAATTAACGAGGGCGCCGACGGA 2862
Qy 221 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIle 240
Db 2863 GTGGGTAAATTCCTCCGGAATTTGGCAATTCGCAATTCACATGGTGGCGACAGTCAATC 2922
Qy 241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIle 260
Db 2923 ACCACGACGACCGGAACCTGGGCGCTGCCCACTACACCAACCACTCTACAAACAAT 2982
Qy 261 SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrp 280
Db 2983 TCCAGCCAAATCA--GGAGCCTCGAACGACAACTACTTTGGCTTACAGCACCCCTTGG 3039
Qy 281 GlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu 300
Db 3040 GGGTATTTTGGACTTCAACAGATTCCACTGCCACTTTTACCACGCTGACTGGCAAGACTC 3099
Qy 301 IleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGln 320
Db 3100 ATCAACAACAATCGGGATTCGACCCCAAGAGACTCAACTTCAAGCTCTTTAACTTCAA 3159
Qy 321 ValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThr 340
Db 3160 GTCAAAAGAGTCAAGCAGAAATGACGGTACGACGAGATTCGCAATTAACCTTACCAACG 3219
Qy 341 ValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGln 360
Db 3220 GTTCAGGTGTTACTGACTCGGAGTACCAGTCCCGTAGCTCTCGGCTCGGCGCATCAA 3279
Qy 361 GlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr 380
Db 3280 GGATGCTCCGCGCGTTCCAGCAGACGCTTTCATGTTGCCACAGTATGGATACCTCAC 3339
Qy 381 LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 400
Db 3340 CTGAACAACGGGAGTCAGGACGATGAGCGCTCTTCAATTTTACTGCGCTGGAGTACTTCT 3399
Qy 401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 420
Db 3400 TCTCAGATGCTGGTACCGAACAACACTTTTACTCTTACGCTACACTTTTGGAGCGTCTCT 3459
Qy 421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
Db 3460 TTCCACAGCAGCTACGCTCACAGCCAGAGTCTGGACCGTCTCATGAATCTCTCATCGAC 3519
Qy 441 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAsp 460
Db 3520 CAGTACCTGTATTTACTTGACAGACAACAACACTTCCAAAGTGGAGAACCCACCGACGCA 3579
Qy 461 LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuPro 480
Db 3580 CTTTCAGTTTCTCAGGCCGAGGAGTGACATTCGGGACCACTAGGAACTGGCTTCTCT 3639
Qy 481 GlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsn 500
Db 3640 GGACCTGTTTACCGCAGCAGCGAGTATCAAGACATCTCGCGGATTAACAACACAGTGAA 3699

Qy 501 PheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnPro 520
Db 3700 TACTCTGGAGCTGGAGCTACCAAGTACCACTCAATGGCAGAGACTCTCTCTGTAATCCG 3759
Qy 521 GlyThrAlaMetAlaSerHisLysAspGluAspLysPhePheProMetSerGlyVal 540
Db 3760 GGGCCCGCATGGACGCCACAGGACGATGAGAAAGATTTTTTCTCAGAGCGGGTT 3819
Qy 541 MetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIle 560
Db 3820 CTCATCTTTGGAGCAAGGCTCAGAGAAAACAAATGTGAACATTTGAAAAGGTCAATGAT 3879
Qy 561 ThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrVal 580
Db 3880 ACAGACGAAGAGAAATCGGAACCAATCCCGTGGCTACGGAGCAGATGGTGTCTGTA 3939
Qy 581 AlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGly 600
Db 3940 TCTACCACTCTCAGAGAGGCAACAGACAGCTACCGCAGATGTCAACACACAAGGC 3999
Qy 601 AlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlnGlyProIleTrpAla 620
Db 4000 GTTCTTCCAGCATGGTCTGGCAGCAGAGATGTGTACCTTCAGGGGCCCATCTGGGCA 4059
Qy 621 LysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeu 640
Db 4060 AAGATTCCACACACGGACGGACATTTTACCCCTCTCCCTCATGGGGTGGATTCGGACT 4119
Qy 641 LysAsnProProGlnIleLysAsnThrProValProAlaAsnProProAla 660
Db 4120 AAACACCTCTCTCCACAGATTTCTCATCAAGAACACCCCGGTACCTGCGAATCTTCGACC 4179
Qy 661 GluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValIse 680
Db 4180 ACCITTCAGTGGGCAAGTTTGTCTTCTCATCACAGTAGTACTCCACGGCACCGGTGAG 4239
Qy 680 rValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValG1 700
Db 4240 CGTGAGATCAGTGGGAGCTGCAAGAGAAACAGCAACGCTGGATCCGGAATTCAC 4299
Qy 700 nTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnGlyLe 720
Db 4300 GTACACTTCCAACATACAAAGTCTGTTAATCGTGGACTTACCGTGGATACTAATGGCGT 4359
Qy 720 uTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 4360 GTATTTCAGAGCTCGCCCATTTGGCACCAGATACCTGACTCGTAATCTG 4408

RESULT 8

PCT-US95-07178-1

; Sequence 1, Application PC/TUS9507178

; GENERAL INFORMATION:

; APPLICANT: Johnson, Philip R.

; TITLE OF INVENTION: Adeno-Associated Virus Materials and

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

; STREET: 6300 Sears Tower, 233 S. Wacker Drive

; CITY: Chicago

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/07178

; FILING DATE:

; CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Noland, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31975
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4680 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US95-07178-1

Alignment Scores:
Pred. No.: 9,71e-315 Length: 4680
Score: 3360.50 Matches: 609
Percent Similarity: 89.42% Conservative: 50
Best Local Similarity: 82.63% Mismatches: 76
Query Match: 84.24% Indels: 2
DB: 5 Gaps: 1

US-09-807-802A-13 (1-736) x PCT-US95-07178-1 (1-4680)

Qy 1 MetAlaAlaAspGlyTyrLeuProAspTyrLeuGluAspLeuSerGluGlyLeuArg 20
Db 2203 ATGCTCCGATGTTATCTTCAGATTGGCTCAGAGCACTCTCTCTGAAGGANTAGA 2262

Qy 21 GluTrpTrpAspLeuLeuProGlyAlaProLysProLysAlaAsnGlnGlnAsp 40
Db 2263 CAGTGGTGAAGCTCAAACTGGCCACCACCAAGCCCGCAGCGGCAATAGGAC 2322

Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLeuGlyTyrLeuGlyProPheAsnGlyLeuAsp 60
Db 2323 GACAGCAGGGGTCTTGTCTCTCTGGGTACAGTACCTCGGACCTTCAACGGACTCGAC 2382

Qy 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
Db 2383 AAGGGAGAGCGGTCTACAGCGCAGACCGCCGGCCCTCTGAGCAGCAACAAAGCTTACGAC 2442

Qy 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
Db 2443 CGCAGCTCGACAGCGGAGACACCGGTACCTCAAGTACAAACCAACCGCGCGGAGTTT 2502

Qy 101 GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln 120
Db 2503 CAGGAGCGCCCTTAAGAAGATACGTCTTTTGGGGCAACCTCGGACGAGCTCTTCCAG 2562

Qy 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGlyAlaLysThrAlaPro 140
Db 2563 CGGAAAAGAGGGTCTTGAACCTCTGGCCCTGGTTGAGGAACTGTGTAAGACGGCTCCG 2622

Qy 141 GlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerGlyLeuGly 160
Db 2623 GGAAGAAAGCGCGGTAGAGCACTCTCTCTGGAGCCAGACTCTCTCTCGGGAACCGGA 2682

Qy 161 LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
Db 2683 AAGCGGCCAGCAGCGCTTGAAGAAAGATGAAATTTTGTGTGAGCTGGAGAGCGCAGAC 2742

Qy 181 SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro 200
Db 2743 TCAGTACCTGACCCCGACCTCTCGGACAGCACCAACCGCCCTCTGTGTCTGGGAAT 2802

Qy 201 ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnGlnGlyAlaAspGly 220
Db 2803 AATACGATGGTACAGGAGGTGGCGCAATATGGCAGCAATATGAGGGCGCGCGCA 2862

Qy 221 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValile 240
Db 2211 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValile 240

Db 2863 GTGGGTAAATTTCTCCGAAATTTGGCAATTTCCATGATGGCGACAGAGTCATC 2922

Qy 241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnile 260
Db 2923 ACCACCGACACCGAAGCTTGGGCTTGGCCCTTACCAACCAACCACTCTACAAACAAT 2982

Qy 261 SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTyr 280
Db 2983 TCCAGGCAATCA---GGAGCTCGAAGCAATCACTACTTTTGGCTACAGCACCCCTTGG 3039

Qy 281 GlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTyrGlnArgLeu 300
Db 3040 GGGTATTTTGACTTCAACAGATTCCACTTCCACTTTTCCACGTCGACTGGCAAGACT 3099

Qy 301 IleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGln 320
Db 3100 ATCAACAACTCTGGGATTCGACCCNAGAGACTCAACTTCAAGCTCTTTAAACATTCAA 3159

Qy 321 ValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnLeuLeuThrSerThr 340
Db 3160 GTCAAGAGAGGTCAACGCAAGATGACGGTACGACGACGATTGCCAATAACCTTACCAGCAG 3219

Qy 341 ValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGln 360
Db 3220 GTTCAGGTGTACTGACTCGGAGTACCAGCTCCCGTACGCTCTCGGCTCGGGCATCAA 3279

Qy 361 GlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr 380
Db 3280 GGATGCTCCCGCGTTCACGACGACGCTCTCATGGTGCACAGTATGATGATACCTCACC 3339

Qy 381 LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 400
Db 3340 CTGAACAAACGGGAGTCAGGAGTAGGACGCTCTTCAATTTTACTGCTCGGAGTACTTTCCT 3399

Qy 401 SerGluMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 420
Db 3400 TCTCAGATGCTGGCTACCGGAACAACTTTTACCTTCAGCTACACTTTTGGAGGAGTCTCT 3459

Qy 421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
Db 3460 TTCCACAGCAGCTACGCTCACAGCAGAGTCTGGACGCTCTCATGATCTCTCTCATCGAC 3519

Qy 441 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAsp 460
Db 3520 CAGTACCTGTATTTACTTTGAGCAGACAAACACTCAAGTGGAAACCAACCGCAGCTCAAGG 3579

Qy 461 LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuPro 480
Db 3580 CTTCAGTTTTCTCAGCGCGGAGCGAGTGACATTCGGGACCACTGTAGAACTGGCTTCTCT 3639

Qy 481 GlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsn 500
Db 3640 GGACCTTGTACCGCCAGCAGCAGGATATCAAGACATCTCGGATATCAACACAGTGAA 3699

Qy 501 PheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnPro 520
Db 3700 TACTCGTGGACTGGAGTACCAAGTACCACCTCAATGGCAGAGACTCTCTGTGTAAATCCG 3759

Qy 521 GlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyVal 540
Db 3760 GGGCCGCCATGTCAGACCCACAGGACGATGAAGAAAGTTTCTCTCAGAGCGGGGTT 3819

Qy 541 MetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIle 560
Db 3820 CTCATCTTTGGGAAGCAAGCTCAGAGAAAAACAAATGTGAACATTTGAAAAGGTCTATGATT 3879

Qy 561 ThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrVal 580
Db 3880 ACAGACGAAGAGGAAATCGGAACCAACCAATCCGCTGGCTACGGAGCAGTATGGTCTCTGTA 3939

Qy 581 AlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGly 600
Db 3940 TCTACCAACTCCAGAGAGCAACAGACAGAGCTTACCGCAGATGTCAACACACAAAGGC 3999

```
QY 601 AlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAla 620
Db 4000 GTTCTTCAGGAGTGGTTCGGCAGACAGAGATGTGTACCTTTCAGGGGCCCATCTGGGCA 4059
QY 621 LysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeu 640
Db 4060 AAGATTTCACACAGCGAGCGACATTTTCCACCCCTCTCCCTCATGGGTGGATTCCGACTT 4119
QY 641 LysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAla 660
Db 4120 AAACACCTCTCTCCACAGATTCTCATCAAGAACACCCCGGTACCTGGGAATCCCTTCGACC 4179
QY 661 GluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGln-Valse 680
Db 4180 ACCTTCAGTGGCGCAAGATTTTCTCTTCATCATCACAGTACTCCACGGGACACGGTCAG 4239
QY 680 rValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValG1 700
Db 4240 CGTGAGATCGAGTGGGAGCTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4299
QY 700 nTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLe 720
Db 4300 GTACACTTCCAACTACAAACAAGTCTGTTAATCGTGGACTTACCGTGGATACTAATGGCGT 4359
QY 720 uTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 4360 GTATTTCAGAGCTCGCCCCATTTGGCACCAGATACCTGTGATCTGTAATCTG 4408

RESULT 9
US-09-438-268-2
; Sequence 2, Application US/09438268
; Patent No. 6491907
; GENERAL INFORMATION:
; APPLICANT: Rabinowitz, Joseph E.
; APPLICANT: Samulski, Richard J
; APPLICANT: Xiao, Weidong
; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 5470-186
; CURRENT APPLICATION NUMBER: US/09438,268
; CURRENT FILING DATE: 1999-11-10
; EARLIER APPLICATION NUMBER: 60/107,840
; EARLIER FILING DATE: 1998-11-10
; EARLIER APPLICATION NUMBER: 60/123,651
; EARLIER FILING DATE: 1999-03-10
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 8151
; TYPE: DNA
; ORGANISM: Virus
US-09-438-268-2

Alignment Scores:
Pred. No.: 3 67e-283 Length: 8151
Score: 3038.50 Matches: 552
Percent Similarity: 83.85% Conservatives: 66
Best Local Similarity: 74.90% Mismatches: 108
Query Match: 76.17% Indels: 11
DB: 4 Gaps: 3

US-09-807-802A-13 (1-736) x US-09-438-268-2 (1-8151)

QY 1 MetAlaAlaAspGlyTyrLeuProAspTrpLeuGluAspAsnLeuSerGluGlyIleArg 20
Db 2133 ATGGCTGGCGATGGTATCTTCAGATTGGCTCGAGGACACTCTCTCTCAAGGAATAAGA 2192
QY 21 GluTrpTrpAspLeuLeuProGlyAlaProLysProLysAlaAsnGlnLysGlnAsp 40
Db 2193 CAGTGGTGAAGCTCAAACTCGGCCACCACCACCAAGAGCCGCGAGCGGATGAAGGAC 2252
QY 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
```

```
Db 2253 GACAGAGGGGTCTTGTGCTTCTCGGTACAGTACTCTCGACCCCTTCAACGGAGCTCGAC 2312
QY 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
Db 2313 AAGGAGAGCGGTCAACGAGGCGAGACGCGCGCCCTCGAGCACGACAAAGGCTACGAC 2372
QY 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
Db 2373 CACAGCTCAAGCCCGGTGACAAACCTTACCTCAAGTACAAACACGCGCGAGGAGTTC 2432
QY 101 GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln 120
Db 2433 CACAGCGGCTTCAGCGGACACATCGTTTGGGGGACCTCGGACAGAGAGTCTTCAG 2492
QY 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaPro 140
Db 2493 GCCAAAAGAGGGTCTTGAACCTCTTGGTCTGGTTCGAGCAAGCGGTGAGACGGCTCT 2552
QY 141 GlyLysLysArgProValGlnSerProGlnGluProAspSerSerSerGlyIleGly 160
Db 2553 GGAAGAGAGAGACCGTTGATTGAATCCCCCAGAGCCCGACTCTCTCCACGGGTATCGGC 2612
QY 161 LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPhe---GlyGlnThrGlyAspSer 179
Db 2613 AAAAAGGCAAGCAGCGCGCTAAAGAGAGCTCGTTTTCGAAGACGAAACTGGAGCAGGC 2672
QY 180 GluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGly 199
Db 2673 GACGGACCCCTGAGGATCAACTTCGGA-----GCCATGTCTGTAT 2714
QY 200 ProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAsp 219
Db 2715 GACAGTGAGATGGCTGCAGCAGCTGGCGAGCTGCAGTCGAGCGGCGGCAAGGTGCCGAT 2774
QY 220 GlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgVal 239
Db 2775 GGAGTGGTAAATGCTTCGGGTGATTGGCATTTGCCATTCACCTGGTCTGAGGGGCCAGTC 2834
QY 240 IleThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGln 259
Db 2835 ACGACCACAGCAGCACCAACCTGGGTCTTGGCCACCTACAAACCAACCTCTACAAACGCA 2894
QY 260 IleSerSerAlaSerThrGlyAlaSerAsnAsnHisTyrPheGlyTyrSerThrPro 279
Db 2895 CTCGGAGAG-----AGCTGCAGTCCACACCTTACACGGATCTCCACCCCC 2942
QY 280 TrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArg 299
Db 2943 TGGGATATCTTGCATTTCAACCGCTTCCACTGCCACTTCTCACCCAGTACTGGCAGCGA 3002
QY 300 LeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIle 319
Db 3003 CTCATCAACCAACATCGGGCATCGCACCACCAAGCCATCGGGTCAAAATCTTCAACATC 3062
QY 320 GlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSer 339
Db 3063 CAGGTCAAGGAGGTTCACAGCTCGAAGCGGAGACACGGTGGCTAATAACCTTACCAGC 3122
QY 340 ThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHis 359
Db 3123 ACGGTTTCAGATCTTTGGGACTCTGTCGTAGCAACTGCCGTACGCTCTCGGCTCGGCGCAT 3182
QY 360 GlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeu 379
Db 3183 CAAGGATGCTTCCCGCGTTCCTCCAGCAGAGAGCTTTCATGTGGCCACAGATATGGATACCTC 3242
QY 380 ThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhe 399
Db 3243 ACCCTGAACACCGGAGTCAGGAGTAGGAGCTCTTCAATTTACTGCTCGAGTACTTT 3302
QY 400 ProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluVal 419
```

Db 3303 CTTCTCAGATGTCGCTACCGAAACAACTTTACCTTCAGCTACACCTTTTGAGGACGTT 3362
Qy 420 PropheHisSerSerTyAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIle 439
Db 3363 CTTTTCACAGAGCTAGCTCAGACGAGAGTCTGGACCGTCTCATGAATCTCTCATC 3422
Qy 440 AspGlnTyLeuTyTyLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLys 459
Db 3423 GACCAGTACCTGTATTACTTGAGCAGACAAACAACTCCCAAGTGGAAACCAACGACGTC 3482
Qy 460 AspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeu 479
Db 3483 AGGCTTCAGTTTCTCAGCGCGGAGCTGACATTCGGGACCACTCTAGGAATCTGGCTT 3542
Qy 480 ProGlyProCysTyArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSer 499
Db 3543 CTTGACCTCTGTTACCGCAGCAGCGAGTATCAAGACATCTCGCGATPACAAACACAGT 3602
Qy 500 AsnPheThrTrpThrGlyAlaSerLysTyAsnLeuAsnGlyArgGluSerIleIleAsn 519
Db 3603 GAATACTCGTGGAGCTACCAAGTACCACCTCACTCAATGGCAGAGACTCTCTGGTGAAT 3662
Qy 520 ProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGly 539
Db 3663 CCGGCGCGGCGCATGGCAAGCCACCAAGGACGATGAAGAAAGTTTTTCTCAGAGCGG 3722
Qy 540 ValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMet 559
Db 3723 GTTCTCATCTTTGGAGACAGGCTCAGAGAAACAAATGTGAACATTTGAAGAGTCAATG 3782
Qy 560 IleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThr 579
Db 3783 ATTACAGACGAGAGAAATCGGAACAACTCCCGTGGCTACGAGCAGTATGTTCT 3842
Qy 580 ValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMet 599
Db 3843 GTATCTACCACTTCAGAGAGGCAACAGACAGCAGCTACCGCAGATGTCAACACACAA 3902
Qy 600 GlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyLeuGlnGlyProIleTrp 619
Db 3903 GCGTTCCTCCAGGATGGTCTGGCAGCAGAGATGTACCTTCAGGGGCCCATCTGG 3962
Qy 620 AlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGly 639
Db 3963 GCAAGATTCACACACGACGACATTTTCAACCCCTCTCCCTCATGGTGGATTGGA 4022
Qy 640 LeuLysAsnProProGlnIleLeuLysAsnThrProValProAlaAsnProPro 659
Db 4023 CTTAAACACCTCTCCACAGATTCCTCATCAAGAACACCCCGGTACCTGCGAATCTTCG 4082
Qy 660 AlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTySerThrGlyGlnVal 679
Db 4083 ACCACTTCAGTCGGGAAAGTTTCTCTTCATCACACAGTATCTCCAGCGGACAGTCTC 4142
Qy 680 SerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluVal 699
Db 4143 AGCGTGAGATCGAGTGGGAGCTGAGAAGGAAGAAACAGCAACGCTGGAATCCCGAAT 4202
Qy 700 GlnTyThrSerAsnTyAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGly 719
Db 4203 CAGTACACTTCCAACTACAACAAAGTCTGTATATCGTGGACCTTACCGTGGATTAATGGC 4262
Qy 720 LeuTyThrGluProArgProIleGlyThrArgTyLeuThrArgProLeu 736
Db 4263 GTGTATTGAGACCTCGCCCCATTGGCCAGATACCTGACTCGTAATCTG 4313

RESULT 10

US-09-532-594B-5

; Sequence 5, Application US/09532594B

; Patent No. 6468524

; GENERAL INFORMATION:

; APPLICANT: Chorini, John A.

; APPLICANT: Kotin, Robert M.

; APPLICANT: Safer, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAVA VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.025202
; CURRENT APPLICATION NUMBER: US/09/532.594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 2208
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; OTHER INFORMATION: synthetic construct
; NAME/KEY: misc_feature
; LOCATION: 750
; OTHER INFORMATION: n = g, a, c or t(u)
; NAME/KEY: misc_feature
; OTHER INFORMATION: AAVA caspid protein VP1 gene
US-09-532-594B-5

Alignment Scores:
Pred. No.: 1.17e-230 Length: 2208
Score: 2486.50 Matches: 475
Percent Similarity: 74.33% Conservative: 81
Best Local Similarity: 63.50% Mismatches: 161
Query Match: 62.33% Indels: 31
DB: 4 Gaps: 10

US-09-807-802A-13 (1-736) x US-09-532-594B-5 (1-2208)

Qy 4 AspGlyTyTrpLeuProAspTrpLeuGluAspAsnLeuSerGluGlyIleArgGluTrp 23
Db 7 GACGTTTACCTTCAGATTTGGCTAGAGACACACTCTCTGAGAGCGTTCGAGGTGG 66
Qy 24 AspLeuLysProGlyAlaProLysProLysAlaAsnGlnLysGlnAspAspGlyArg 43
Db 67 GCGTGCACCTGGAGCCCTTAACCCAGGCAATCAACATCAGGACCAACGCTCGG 126
Qy 44 GlyLeuValLeuProGlyTyTrpLysTyTrpLysTyTrpLysTyTrpLysTyTrp 63
Db 127 GGTCTTGCTTCCGGGTTTCAATATCTCGGACCCCGGCAACGAGCTCGACAAGGGGAA 186
Qy 64 ProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyAspGlnGlnLeu 83
Db 187 CCGCTCAACGCGGAGCGGCGGCGGCGCTCGAGCACGACGAGGCTACGACGAGCTC 246
Qy 84 LysAlaGlyAspAsnProTyTrpLeuArgTyTrpAsnHisAlaAspAlaGluPheGlnGluArg 103
Db 247 AAGCGCGTGACAACTTACCTCAAGTACCAACCCAGCAGCGGAGTTCCAGCAGCGG 306
Qy 104 LeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGlnAlaLys 123
Db 307 CTTCAAGCGGACACATCGTTTGGGGCAACCTCGCAGAGCAGCTTTCAGCGCCAAAG 366
Qy 124 ArgValLeuGluProLeuGlyLeuValGluGlyAlaLysThrAlaProGlyLysLys 143
Db 367 AGGTTCTTGACCTCTTGTTGGTTGAGCAAGCGGGTGAGCGGCTCTCGGAAGAG 426
Qy 144 ArgProValGluGlnSerProGlnGluProAspSerSerSerGlyIleGlyLysThrGly 163
Db 427 AGACCGTTGATTGAATCCCCCAGCAGCGGCTCTCCAGCGGTATCGGCAAAAAGGC 486
Qy 164 GlnGlnProAlaLysLysArgLeuAsnPhe---GlyGlnThrGlyAspSerGluSerVal 182
Db 487 AAGCAGCGCGCTAAAGAGAGCTCGTTTTCGAGAGCAAACTGGAGCGGCGGAGCC 546
Qy 183 ProAspProGlnProLeuGluProProAlaThrProAlaAlaValAlaGlyProThr 202
Db 547 CCTGAGGATCACTTCGGA-----GCCATGTCTGATGACAGTGA 588
Qy 203 MetAlaSerGlyGlyAlaProMetAlaAspAsnGluGlyAlaAspGlyValGly 222

Db 589 ATGCGTGCAGAGCTGGCGAGCTGCAGTCGAGGSGGCAAGGTGCCGATGAGTGGGT 648
Qy 223 AenAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 242
Db 649 AATGCTCGGGTGTGGGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 708
Qy 243 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrIleGlnIleSerSer 262
Db 709 AGCACCAGAACCTGGGTCTTGCCCACTTACACACACCTTACACACCTTACACACCTT 768
Qy 263 AlaSerThrGlyAlaSerAsnAsnHisLeuTyrIleGlnIleSerSerThrTrpGlyTyr 282
Db 769 -----AGCTGCAGTCCACACCTACACCGGATTCCTCCACCCCTCGGGATAC 816
Qy 283 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 302
Db 817 TTGACTTCAACCGCTTCCACTGCCACTTCTCACCACGTGACTGGCGGACTCATCAAC 876
Qy 303 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLeuPheAsnIleGlnValLys 322
Db 877 AACACTGGGCGATGCGACCCCAAGCCATGCGGGTCAAAATCTTCAACATCCAGGTCAAG 936
Qy 323 GluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeuThrSerThrValGln 342
Db 937 GAGTCCACGCTCGAAGCGGAGACACCGTGGCTAATACTTACCAGCACGGTTCCAG 996
Qy 343 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 362
Db 997 ATCTTTGGGAGCTCGTGCAGCAACTGCGGTGATGATGATGATGATGATGATGATGAT 1056
Qy 363 LeuProPheProAlaAspValPheMetIleProGlnTyrGlyTyr-----Leu 379
Db 1057 CTCCTCTCTTTCCCAACGACGCTTTATGTTGGTCCCGGATGATGATGATGATGATGAT 1116
Qy 380 ThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhe 399
Db 1117 ACCGGCAACACTTCGCGAGCAACAGACTGCAGAGAAATGCTTCTACTGCTGGAGTACTTT 1176
Qy 400 ProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluVal 419
Db 1177 CCTTCGAGATGCTGGGAGCTGGCAACAACTTTGAAATACGATACAGTGTGAGAGGTG 1236
Qy 420 ProPheHisSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIle 439
Db 1237 CCTTTCCACTCGATGACGGCGACAGCCAGAGCTGGACCGGCTGATGAACTCTCATC 1296
Qy 440 AspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsn-----GlnSerGlySer 455
Db 1297 GACCAGTACCTGTGGGAGCTGCAATCGACCACCGCAACCCCTGAATGCGGAGCT 1356
Qy 456 AlaGlnAsnLysAspLeuPheSerArgGlySerProAlaGlyMetSerValGlnPro 475
Db 1357 GCCACCACCAAC-----TTTACCAAGCTCGGCTTACCACCTTTTCCAACTTTAA 1407
Qy 476 LysAsnTrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAsp 495
Db 1408 AAGACTGGTGTCCCGGCTTCAATCAACGACGAGGCTTCTCAAGACTGCC----- 1461
Qy 496 AsnAsnAsnSerAsnPheThrTrpThrGlyAlaSer-----LysTyr----- 509
Db 1462 AATCAAACTACAAAGATCCCTGCGCCAGCGGTGACAGAGTCTCATCAATACGAGCGCAC 1521
Qy 510 ---AsnLeuAsnGlyArgGluSerIleLeuAsnProGlyThrAlaMetAlaSerHisLys 528
Db 1522 AGCACTCTGAGGAGATGAGTGGCTTACCCCGGAGCTTCAATGGCCAGCGGTGGA 1581
Qy 529 AspAspGluAspLysPhePheProMetSerGlyValMetIlePheGlyLysGluSerAla 548
Db 1582 CCTGGGACAGCAAG---TTCAAGCAACAGCAGCTCATCTTTGGGGGCTTAACAGAAC 1638
Qy 549 GlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAspGluGluIleLysAla 568

Db 1639 GGCAACAGCGCCACCGTACCCGGGACTGTGATCTTCACTCTGAGGAGGAGCTGGCAGCC 1698
Qy 569 ThrAsnProValAlaThrGluArgPheGlyThrValAlaValAsnPheGlnSerSerSer 588
Db 1699 ACCAAGCCACCGATACCGACATGTGGGCAACCTACTCTGGCGGTGACACGAGCAACAGC 1758
Qy 589 ThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeuProGlyMetValTrpGln 608
Db 1759 AACCTGGCGCGGTGGACAGACTGACAGCTTGGGAGCGGTGCTGGATGGTCTGGCAA 1818
Qy 609 AspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIleProHisThrAspGlyHis 628
Db 1819 AACAGAGACATTTACTACAGGGTCCCATTGGGCAAGATTCTCTCATACGATGGACAC 1878
Qy 629 PheHisProSerProLeuMetGlyPheGlyLeuLysAsnProProProGlnIleLeu 648
Db 1879 TTTCAACCTCCTCACCGCTGATTTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 1938
Qy 649 IleLysAsnThrProValProAlaAsnProProAlaGluPheSerAlaThrLysPheAla 668
Db 1939 ATCAAGACACCCCGGTACTCTGCGAATCTCTCAACGACCTTCTAGCTTCTCCGTAAC 1998
Qy 669 SerPheIleThrGlnTyrSerThrGlyGlnValSerValGluIleGluTrpGluLeuGln 688
Db 1999 TCCTTTACTTACTCAGTACAGCACTGGCCAGGTGCGTGGTGGTGGTGGTGGTGGTGG 2058
Qy 689 LysGluAsnSerLysArgTyrAsnProGluValGlnTyrThrSerAsnTyrAlaLysSer 708
Db 2059 AAGGAGCGGTCCAAACCGCTGGAACCCCGAGGTCCAGTTTACTCTCACTACGACAGCA 2118
Qy 709 AlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThrGluProArgProIleGly 728
Db 2119 AACTCTCTGTTGGGCTCCCGATCGCGCTGGGAAATACACTGAGCCTAGGCTATCGT 2178
Qy 729 ThrArgTyrLeuThrArgProLeu 736
Db 2179 ACCCGCTACCTCACCCACCCTG 2202

RESULT 11
US-09-438-268-1/c
; Sequence 1, Application US/09438268
; Patent No. 6491907
; GENERAL INFORMATION:
; APPLICANT: Rabinowitz, Joseph E.
; APPLICANT: Samuleki, Richard J.
; APPLICANT: Xiao, Weidong
; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING
; FILE REFERENCE: 5470-186
; CURRENT APPLICATION NUMBER: US/09/438,268
; CURRENT FILING DATE: 1999-11-10
; EARLIER APPLICATION NUMBER: 60/107,840
; EARLIER FILING DATE: 1998-11-10
; EARLIER APPLICATION NUMBER: 60/123,651
; EARLIER FILING DATE: 1999-03-10
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 7214
; TYPE: DNA
; ORGANISM: Virus
US-09-438-268-1

Alignment Scores:
Pred. No.: 8,348-230 Length: 7214
Score: 2486.50 Matches: 475
Percent Similarity: 74.33% Conservatives: 81
Best Local Similarity: 63.50% Mismatches: 161
Query Match: 62.33% Indels: 31
DB: 4 Gaps: 10

US-09-807-802A-13 (1-736) x US-09-438-268-1 (1-7214)

QY 4 AspGlyTyrLeuProAspTyrLeuGluAspAsnLeuSerGluGlyIleArgGluTyrTrp 23
DB 2977 GACGTTACCTTCAGATTGGCTAGAGGACAACTCTCTGAGGCGTTCAGAGTGGTG 2918
QY 24 AsnLeuLysProGlyAlaProLysProLysAlaAsnGlnGlnLysGlnAspGlyArg 43
DB 2917 GCGCTGCAACCTGGAGCCCTAAACCCCAAGGCAAAATCAACAACATCAGGCAACGCTCGG 2858
QY 44 GlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAspLysGlyGlu 63
DB 2857 GGTCTTGCTTCGGGTTCAAATATCTCGGACCCGCGCAACGAGCTCGCAAGGGGAA 2798
QY 64 ProValAsnAlaAspAlaAlaLeuGluHisAspLysAlaTyrAspGlnGlnLeu 83
DB 2797 CCCGTCAACGCGGACGCGGAGCCCTCGAGCACGCAAGGCTTACGACCGAGCTC 2738
QY 84 LysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPheGlnGluArg 103
DB 2737 AAGCCCGGTGCAACCCCTACTCAAGTACAACCAACGCGGAGGATTCAGAGCGG 2678
QY 104 LeuGlnGluAspThrSerPheGlyAsnLeuGlyArgAlaValPheGlnAlaLysLys 123
DB 2677 CTTGAGGCGGACATCGTTTGGGGGCAACCTCGGAGAGCGATCTTCAGGCGCAAAAG 2618
QY 124 ArgValLeuGluProLeuGlyLeuValGluGlyAlaLysThrAlaProGlyLysLys 143
DB 2617 AGGTTCTTGAACTCTTGTTGTTGTTGAGCAAGCGGTGAGACGGTCTCTGGAAGAAG 2558
QY 144 ArgProValGluGlnSerProGlnGluProAspSerSerGlyIleGlyLysThrGly 163
DB 2557 AGACCGTTGATTGAATCCCCCGACGAGCCGACTCTCCAGCGGTATCGGCMAAAAGGC 2498
QY 164 GlnGlnProAlaLysLysArgLeuAsnPhe---GlyGlnThrGlyAspSerGluSerVal 182
DB 2497 ANGACCGCGGTAAAGAGACTCGTTTGAAGACGAACCTGGAGAGCGGCGGACCC 2438
QY 183 ProAspProGlnProLeuGlyGluProProAlaThrProAlaLysValGlyProThrThr 202
DB 2437 CTTGAGGGATCAACTTCGGA-----GCCATGTCTGATGACAGTGAG 2396
QY 203 MetAlaSerGlyGlyGlyAlaProMetAlaAspAsnGlnGlyAlaAspGlyValGly 222
DB 2395 ATGCGTGACAGCTGCGGAGCTGCAGTCGAGGCGGACAAAGTGGCGATGGAGTGGGT 2336
QY 223 AsnAlaSerGlyAsnTyrHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 242
DB 2335 AATGCTCGGTGATGGCATTTGGATTCACCTGCTGAGGGCCAGCTCACGACCACC 2276
QY 243 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnLeuSerSer 262
DB 2275 AGCACCAGAACCTGGGTCTTGCCCACTTACCAACACCACTCTACAAAGCGACTCGGAGAG 2216
QY 263 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 282
DB 2215 -----AGCCTCGAGTCCCAACACCTTACAAAGCGATCTCCACCCCTCGGGGATAC 2168
QY 283 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuLeuAsn 302
DB 2167 TTGACTTCACCGCTTCACCTGCGCATCTTCACCGCTGATCTGAGCGGAGCTCATCAAC 2108
QY 303 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 322
DB 2107 AACAACTGGGCGATCGGACCCCAAGCCATGCGGGTCAAAATCTTCAACATCCAGGTCAAG 2048
QY 323 GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 342
DB 2047 GAGGTCAAGCGTTCGAACGCGGAGCAACCGGTGGCTAATAACCTTACCAGCACCGTTCCAG 1988
QY 343 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 362
DB 1987 ATCTTTGGGACTCTCTGATAGAACTGCGTACGTGATGGATGCGGGTCAAGAGGGGAGC 1928
QY 363 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyr-----Leu 379

DB 1927 CTGCTCTCTTTCCCAACGACGCTCTTATGGTGCCCGCTACGCTACTGTGGAGTGGTG 1868
QY 380 ThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhe 399
DB 1867 ACCGGCAACACTTCGCGACCAACAGACTGACAGAAATGCTCTTACTGCTGGAGTACTTT 1808
QY 400 ProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluVal 419
DB 1807 CTTTGCAGATGCTGGGACTGGCAACAACTTTGAAATACGTACAGTTTGGAGAGGTG 1748
QY 420 ProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIle 439
DB 1747 CTTTCCACTCGATGTACGCGACAGCCAGAGCGCTGGACCGGTGATGAACCCCTCTCATC 1688
QY 440 AspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsn-----GlnSerGlySer 455
DB 1687 GACCACTGCTGGGGAGCTGCAATCGACCAACCCCGGACCACTGAATGATCGGGAGCT 1628
QY 456 AlaGlnAsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnPro 475
DB 1627 GCCACCAACCAAC-----TTTACCAAGCTGCGGCTTACCAACTTTTCCAACATTAAA 1577
QY 476 LysAsnTrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAsp 495
DB 1576 AAGAACTGGCTGCGCGGCTTCAATCAAGCAGCAGGGCTTCTCAAGAGCTGCC----- 1523
QY 496 AsnAsnAsnSerAsnPheThrTrpThrGlyAlaSer-----LysTyr----- 509
DB 1522 AATCAAAACTTACAAGATCCCTGCCCGGCTGACAGCTCATCAATACGAGACGCAC 1463
QY 510 ---AsnLeuAsnGlyArgGluSerIleIleAsnProGlyThrAlaMetAlaSerHisLys 528
DB 1462 AGCACTCTGGACGGAAGATGGAGTGCCTGACCCCGGACCTCCAAATGGCCACGGCTGGA 1403
QY 529 AspAspGluAspLysPhePheProMetSerGlyValMetIlePheGlyLysGluSerAla 548
DB 1402 CTTGCGGACAGCAAG---TTACAGCAACAGCAGCTCATCTTTCGGGGCTTAAACAGAAC 1346
QY 549 GlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAspGluGluGluIleLysAla 568
DB 1345 GGCACACGCGCACCGTACCCGGAGCTGTGATCTTCACTCTGAGGAGGAGCTGGCAGCC 1286
QY 569 ThrAsnProValAlaThrGluArgPheGlyThrValAlaValAsnPheGlnSerSerSer 588
DB 1285 ACCACGCCACCGATACGAGACATGTGGGGCAACCTACTCGCGGTGACAGAGCAACAGC 1226
QY 589 ThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeuProGlyMetValTrpGln 608
DB 1225 AACCTGCGGACCGTGGACAGACTGACAGCCCTTGGAGCGCTGCGTAATGGTGTGGCAA 1166
QY 609 AspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIleProHisThrAspGlyHis 628
DB 1165 AACAGAGACATTTACTACAGGTCCTCATTTGGCCCAAGATTTCTCATACGATGGACAC 1106
QY 629 PheHisProSerProLeuMetGlyPheGlyLeuLysAsnProProGlnIleLeu 648
DB 1105 TTTCAACCCCTCACCGCTGATTGGTGGTGGTGGCTGAAACACACCGCCCTCTCAAAATTTT 1046
QY 649 IleLysAsnThrProValProAlaAsnProProAlaGluPheSerAlaThrLysPheAla 668
DB 1045 ATCAAGAACACCCCGGTACCTGCGAATCTCTGCAACGACCTTACGTCTTCTCCGGTAAAC 986
QY 669 SerPheIleThrGlnTyrSerThrGlyGlnValSerValGluIleGluTrpGluLeuGln 688
DB 985 TCCTTCAATTACTCATGACAGCACTGGCCAGGTGCTGGTGCAGATTGCTGGAGATCCAG 926
QY 689 LysGluAsnSerLysArgTrpAsnProGluValGlnTyrThrSerAsnTyrAlaLysSer 708
DB 925 AAGGAGCGGTCCAAACGCTGGAACCCCGAGGTCCAGTTTACCTTCAACTCAGGACAGCAA 866
QY 709 AlaAsnValAspPheThrValAspAsnGlyLeuTyrThrGluProArgProIleGly 728

Db 865 AACTCTCTGTGGGCTCCGATCGGCTGGGAAATAACACTGACCTAGGCTATCGGT 806
 Qy 729 ThrArgTyrLeuThrArgProLeu 736
 Db 805 ACCCGCTACCTACCCACCACCTG 782

RESULT 12

US-09-532-594B-1
 ; Sequence 1, Application US/09532594B
 ; Patent No. 6468524
 ; GENERAL INFORMATION:
 ; APPLICANT: Chorini, John A.
 ; APPLICANT: Kotin, Robert M.
 ; APPLICANT: Safar, Brian
 ; APPLICANT: Davidson, Beverly
 ; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
 ; FILE REFERENCE: 14014.0252U2
 ; CURRENT APPLICATION NUMBER: US/09/532,594B
 ; CURRENT FILING DATE: 2000-03-22
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1
 ; LENGTH: 4767
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
 ; OTHER INFORMATION: synthetic construct
 ; NAME/KEY: misc feature
 ; LOCATION: 3009_
 ; OTHER INFORMATION: n = g, a, c or t(u)
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: AAV4 genome
 US-09-532-594B-1

Alignment Scores:
 Pred. No.: 1,28e-229 Length: 4767
 Score: 2481.50 Matches: 474
 Percent Similarity: 74.20% Conservative: 81
 Best Local Similarity: 63.37% Mismatches: 162
 Query Match: 62.21% Indels: 31
 DB: 4 Gaps: 10

US-09-807-802a-13 (1-736) x US-09-532-594B-1 (1-4767)

Qy 4 AspGlyTyrLeuProAspTrpLeuGluAspAsnLeuSerGluGlyIleArgGluTrpTrp 23
 Db 2266 GACGGTTACTTCCAGATTGGCTAGAGGACACCTCTCTGAGGGCTTCGAGAGTGGTG 2325
 Qy 24 AspLeuLysProGlyAlaProLysProLysAlaAsnGlnLysGlnAspAspGlyArg 43
 Db 2326 GCGCTGCAACCTGGAGCCCTTAACCCCAAGGCAATCAACAATCAGGACCAACGCTCGG 2385
 Qy 44 GlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlnLysGlyGlu 63
 Db 2386 GGTCTTGTGCTCCGGTTACAAATACCTCGGACCCGCAACGACTCGCAAGGGGNA 2445
 Qy 64 ProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAspGlnGlnLeu 83
 Db 2446 CCGCTCAACGACGAGCGGCGGAGCCCTCGAGCAGCAACAGGCTTACGACGAGCTC 2505
 Qy 84 LysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPheGlnGluArg 103
 Db 2506 AAGCGCGGTGACAAACCCCTACTCTCAAGTACAAACACGCGCGGAGGTTCCAGCAGCGG 2565
 Qy 104 LeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGlnAlaLysLys 123
 Db 2566 CTTACGGCGGACACACCGTTTGGGGGCAACCTCGGACAGGAGCAGTCTTCCAGGCCAANAAG 2625
 Qy 124 ArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaProGlyLysLys 143
 Db 2626 AGGGTTCTTGAACCTCTTGTGCTGTGGTTGAGCAAGCGGCTGAGACGGCTCTCTGGAAGAAG 2685

Qy 144 ArgProValGluGlnSerProGlnGluProAspSerSerSerGlyIleGlyLysThrGly 163
 Db 2686 AGACCGTTGATTGAATCCCCCAGCAGCCGACTCTCTCCACGGGTATCGCAAAAAAGGC 2745
 Qy 164 GlnGlnProAlaLysValArgLeuAsnPhe--GlyGlnThrGlyAspSerGluSerVal 182
 Db 2746 AGCAGCCGGCTAAAGAGAGCTCGTTTTCGAGACGAAACTGGAGCAGGCGGAGCC 2805
 Qy 183 ProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyProThrThr 202
 Db 2806 CCGTGGGATCAACTTCGGGA-----GCCATGTGTGATGACAGTGAG 2847
 Qy 203 MetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGly 222
 Db 2848 ATGCGTCGACGAGCTGGCGAGCTGCAGTCGAGGGGAGCAAGGTGCGATGGAGTGGGT 2907
 Qy 223 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 242
 Db 2908 ATGCGCTCGGGTGTGGCATTTGGATTCCACCTGGTCTGAGGGCCACGTCACGACCAC 2967
 Qy 243 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSer 262
 Db 2968 AGCACCAGAACCTGGGTCTTGGCCACCTACAAACACCACTTACAAAGCGACTCGAGAG 3027
 Qy 263 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 282
 Db 3028 -----AGCCTGCGAGTCCAACACCTTACACCGGATTTCTCACCCCTCGGGATAC 3075
 Qy 283 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 302
 Db 3076 TTTGACTTCAACCGCTTCCACTGCCACTTCTCACCACTGCTGAGCGGACTCATCAAC 3135
 Qy 303 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 322
 Db 3136 AACAACTGGGCGATGCGACCCCAAGCCATGCGGGTCAAAATCTTCAACATCCAGTCAAG 3195
 Qy 323 GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 342
 Db 3196 GAGTTCAGCAGCTCGAACCGGCGAGACACCGTGGCTTAATACTTACGACGACCGTTGAG 3255
 Qy 343 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 362
 Db 3256 ATCTTTGGGACTCGTGTACGAACCTGCGTACGTGATGGATGCGGGTCAAGAGGCGACG 3315
 Qy 363 LeuProPheProAlaAspValPheMetIleProGlnTyrGlyTyr-----Leu 379
 Db 3316 CTGCTCTCTTTTCCCAACGACGCTCTTATGCTGCCAGTACGCGTACTGTGGAGTGGTG 3375
 Qy 380 ThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhe 399
 Db 3376 ACCGGCAACATTCGACGACACACTGACAGAAATGCTTCTACTGCTGGAGTACTTT 3435
 Qy 400 ProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluVal 419
 Db 3436 CTTTCGACAGTGTGCGGACTGGCAACAACCTTTGAAATTTAGCTACAGTCTTTGAGAAGTG 3495
 Qy 420 PropheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIle 439
 Db 3496 CTTTCCACTCGATGTACGCGCACAGCAGGCGCTGGAGCCCGGCTGATGAACCTCTCATC 3555
 Qy 440 AspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsn-----GlnSerGlySer 455
 Db 3556 GACCACTACTGTGGGAGCTCAATCGACACCCAGCAACCCCTGAATGCCGGGACT 3615
 Qy 456 AlaGlnAsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnPro 475
 Db 3616 GCCACCAACAAC-----TTTACCAAGCTGGCGCTTACCAACTTTTCAACTTTAA 3666
 Qy 476 LysAsnTrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAsp 495
 Db 3667 AAGAACTGGCTGCCGGGCTTCAATCAAGCAGCAGGCGCTTCTCAAGACGCTGCC----- 3720
 Qy 496 AsnAsnAsnSerAsnPheThrTrpThrGlyAlaSer-----LysTyr----- 509

```
Db 3721 AATCAAACTACAAGATCCCTGCCACCGGTCCAGACAGTCTCATCAATAACGAGCGCAC 3780
Qy 510 ---AsnLeuAsnGlyArgGluSerIleIleAsnProGlyThrAlaMetAlaSerHisLys 528
Db 3781 AGCATCTGGAGGGAAGATGGATGCCCTGACCCCGGACCTCCCAATGGCCACGGCTCGA 3840
Qy 529 AspAspGluAspLysPhePheProMetSerGlyValMetIlePheGlyLysGluSerAla 548
Db 3841 CCTGGGAGCAGCAAG---TTACAGCAACAGCCAGCTCATCTTTGGCGGCGCTAAACAGAAC 3897
Qy 549 GlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAspGluGluGluIleLysAla 568
Db 3898 GGCAACACGCGCACCGTACCCCGGACTCTGATCTTCACCTCTGAGGAGGAGGTGGCAGCC 3957
Qy 569 ThrAsnProValAlaThrGluArgPheGlyThrValAlaValAsnPheGlnSerSerSer 588
Db 3958 ACCAAGCCACCGATACGACATGTTGGGCAACCTTACCTGGCGGTGACAGACAGCAC 4017
Qy 589 ThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeuProGlyMetValTrpGln 608
Db 4018 AACCTGCCGACCGTGACAGACTGACAGCCTTGGGAGCGCTGCTGGAATGCTCTGGCAA 4077
Qy 609 AspArgAspValTyrluGlnGlyProIleTrpAlaLysIleProHisThrAspGlyHis 628
Db 4078 AACAGAGACATTTACTACCGAGGTCCCATTTGGGCCAAGATTCTCTATACCGATGGACAC 4137
Qy 629 PheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsnProProGlnIleLeu 648
Db 4138 TTTCACCCCTCACCGCTGATTGGTGGTTGGCTGAAACCCGCCCTCCTCAAAATTTT 4197
Qy 649 IleLysAsnThrProValProAlaAsnProProAlaGluPheSerAlaThrLysPheAla 668
Db 4198 ATCAAGAACACCCCGTACTCGGAATCTGCAACAGCCTTCAGCTCTACTCCGTAAC 4257
Qy 669 SerPheIleThrGlnTyrluSerThrGlyGlnValSerValGluIleGluTrpGluLeuGln 688
Db 4258 TCCCTTCATTTACTGATACAGACTGGCCAGGTGCGGTGCAGATTGACTGGGAGATCCAG 4317
Qy 689 LysGluAsnSerLysArgTrpAsnProGluValGlnTyrluSerAsnTyrluAlaLysSer 708
Db 4318 AAGGAGCGGTCCAAACGCTGGAAACCCCGAGGTCCAGTTTACTTCCAACTACGACAGCAA 4377
Qy 709 AlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrluThrGluProArgProIleGly 728
Db 4378 AACTCTCTGTGTGGCTCCCGATCGCGGTGGGAAATACTAGGCTAGGCTATCGT 4437
Qy 729 ThrArgTyrluThrArgProLeu 736
Db 4438 ACCCGCTACTCTACCCACACACCTG 4461
```

RESULT 13

```
US-09-532-5948-17
; Sequence 17, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chordin, John A.
; APPLICANT: Kotin, Robert M.
; APPLICANT: Safer, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.025202
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1800
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; OTHER INFORMATION: synthetic construct
```

```
; NAME/KEY: misc feature
; LOCATION: 342
; OTHER INFORMATION: n = g, a, c or t (u)
; NAME/KEY: misc feature
; OTHER INFORMATION: AAV4 capsid protein VP2 gene
US-09-532-5948-17
```

```
Alignment Scores:
Pred. No.: 2,72e-167 Length: 1800
Score: 1830.50 Matches: 355
Percent Similarity: 69.87% Conservative: 74
Best Local Similarity: 57.82% Mismatches: 154
Query Match: 45.89% Indels: 31
DB: 4 Gaps: 10
```

US-09-807-802A-13 (1-736) x US-09-532-594B-17 (1-1800)

```
Qy 138 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSer 157
Db 1 ACGGCTCTCTGAAAGAAGAGACCGTTGATTGAATCCCCCGAGCCGACTCTCTCCACG 60
Qy 158 GlyIleGlyLysThrGlyGlnProAlaLysLysArgLeuAsnPhe---GlyGlnThr 176
Db 61 GGTATCGGCAAAAAGGCAAGCAGCCCGCTAAAAGAAGCTCGTTTTCGAAGACGAAC 120
Qy 177 GlyAspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAla 196
Db 121 GGAGCAGCGCAGCGACCCCTGAGGGATCAACTTCCGGA-----GCC 162
Qy 197 AlaValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGlu 216
Db 163 ATGTCGTGATGACATGAGTGCCTGAGCAGCTGCGCGAGCTGCAGTGGGGGGGACAA 222
Qy 217 GlyAlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGly 236
Db 223 GGTCCCATGGAGTGGGTAAATCCCTCGGGTATTGGCATTCGGATTCCACCTGCTCTGAG 282
Qy 237 AspArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTyrluAsnAsnHisLeu 256
Db 283 GGCCACGCTCACGACACACAGCAGCAACCTGGGTCTTGGCCACCTACACACACACCTCN 342
Qy 257 TyrluSerGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrluPheGlyTy 276
Db 343 TACAAGCGACTCGGAGAG-----AGCCTGCAGTCCCAACACCTCAACCGGATTC 390
Qy 277 SerThrProTrpGlyTyrluPheAspPheAsnArgPheHisCysHisPheSerProArgAsp 296
Db 391 TCCACCCCTCGGGGATTTGACTTCAACCGCTTCCACTGCCACTTCTACACCGTAC 450
Qy 297 TrpGlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeu 316
Db 451 TGGCAGCGACTCATCAACCAACTGGGGCATGGACCCAAAGCCATGCGGGTCAAAATC 510
Qy 317 PheAsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsn 336
Db 511 TTCAACATCCAGGTCAAAGGAGGTCAACAGCTCGAACCGCGGAGACAACGGTGGCTAATAAC 570
Qy 337 LeuThrSerThrValGlnValPheSerAspSerGluTyrluGlnLeuProTyrluValLeuGly 356
Db 571 CTTACACGACCGTTTCAGATCTTTGCGGACTCTGTCGTACGAACCTGCCGTACGTATGAT 630
Qy 357 SerAlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrl 376
Db 631 CGGGGTCAAGAGGCGAGCTGCTCTCTTTCCCAACGACGCTCTTTATGTGTGCCCGAC 690
Qy 377 GlyTyrlu-----LeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPhe 393
Db 691 GGCTACTGTGACTGGTGGCGGCAACACTTCCGAGCAACAGACTGCACAGAAATGCCTTC 750
Qy 394 TyrluLeuGluTyrluPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSer 413
Db 751 TACTGCTGGAGTACTTTCTCTCCAGATGCTGGGAGTGGCAACAACTTTTGAAATTACG 810
```

| | | | |
|----|------|---|------|
| Qy | 414 | TyThrPheGluGluValProPheHisSerSerTyAlaHisSerGlnSerLeuAspArg | 433 |
| Db | 811 | TACAGCTTTTGGAGAAAGTGCTCTTCCATCTGATGTACGGCACAGCCAGACCTGGACCGG | 870 |
| Qy | 434 | LeuMetAsnProLeuLeuAspGlnTyLeuTyTyLeuAsnArgThrGlnAsn | 451 |
| Db | 871 | CTGATGAACCTCTCATCGACCAAGTACCTGTGGGACTGCAATCGACCAACCACCGGAACC | 930 |
| Qy | 452 | -----GlnSerGlySerAlaGlnAsnLysAspLeuLeuPheSerArgGlySerProAla | 469 |
| Db | 931 | ACCTGAATCCCGGACTGCGACCAAC-----TTTACCAAGCTGGCGCTTACC | 981 |
| Qy | 470 | GlyMetSerValGlnProLysAsnTrpLeuProGlyProCytyrArgGlnGlnArgVal | 489 |
| Db | 982 | AACCTTTTCCAACTTAAAGAAAGACTGGCTGCCCGGGCTTCATCAAGCAGCAGGCGCTTC | 1041 |
| Qy | 490 | SerLysThrLysThrAspAsnAsnAsnSerAsnPheThrTrpThrGlyAlaSer | 507 |
| Db | 1042 | TCAAAGACTGCC-----AATCAAAATACAGAATCCCTGCCACCGGTGACAGACTCTC | 1095 |
| Qy | 508 | ---LysTyx-----AsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr | 522 |
| Db | 1096 | ATCAATACAGACGCCACAGACTCTGGACGGAGATGGAGTGCTCTGACCCCGGACCT | 1155 |
| Qy | 523 | AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle | 542 |
| Db | 1156 | CCAATGCCACCGCTGGACCTTCGGACGAGCAAG---TTACGACCAACGCCAGCTCATCTTT | 1212 |
| Qy | 543 | PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp | 562 |
| Db | 1213 | CGGGGGCTAAACAGACGGCAACGGCCACCGTACCCGGGACTCGATCTTCACCTCT | 1272 |
| Qy | 563 | GluGluGluLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal | 582 |
| Db | 1273 | GAGGAGAGCTGGCAGCGCACCAACGCCACCGATACGACATGTGGGCAACCTACCTGGC | 1332 |
| Qy | 583 | AsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu | 602 |
| Db | 1333 | GGTCACACAGACCAACGACCTCCAGCGTGGACAGACTGACAGCGCTTGGGAGCGGTG | 1392 |
| Qy | 603 | ProGlyMetValTrpGlnAspArgAspValTyLeuGlnGlyProIleTrpAlaLysIle | 622 |
| Db | 1393 | CCTGGAATGTGCTGGCAAAAACAGAGACATTTACTACGAGGTGCCATTTTGGGCAAGATT | 1452 |
| Qy | 623 | ProHisThrAspGlyHisPheHisProSerProLeuMetGlyLysPheGlyLeuLysAsn | 642 |
| Db | 1453 | CCTCATACCGATGACACTTTCACCCCTCACCGCTGATGGTGGGTGGGGCTGGAACAC | 1512 |
| Qy | 643 | ProProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe | 662 |
| Db | 1513 | CCGGCTCTCAAAATTTTATCAAGAACACCCCGGTAGCTCGCAATCTCGCAAGCACCTTC | 1572 |
| Qy | 663 | SerAlaThrLysPheAlaSerPheIleThrGlnTySerThrGlyGlnValSerValGlu | 682 |
| Db | 1573 | AGCTCTACTCCGGTAAACTCTTCTTACTCAGTACAGCACTGGCCAGGTGTGGTGACG | 1632 |
| Qy | 683 | IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyThr | 702 |
| Db | 1633 | ATTGACTGGGAGATCCAGAAAGAGCGGTCCAAACCGCTGGAACCCCGAGGTCCAGTTTACC | 1692 |
| Qy | 703 | SerAsnTyAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyThr | 722 |
| Db | 1693 | TCCAACTACGACAGCAAAAATCTCTGTGTGGGCTCCCGATGGCGGTGGGAATATCACT | 1752 |
| Qy | 723 | GluProArgProIleGlyThrArgTyLeuThrArgProLeu | 736 |
| Db | 1753 | GAGCTTAGGGTATCGGTACCCGCTTACTCAACCCACCACTGTG | 1794 |

```

; APPLICANT: Chorini, John A.
; APPLICANT: Kotin, Robert M.
; APPLICANT: Safer, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.0252U2
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 1617
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; OTHER INFORMATION: synthetic construct
; NAME/KEY: misc_feature
; LOCATION: 162
; OTHER INFORMATION: n = g, a, c or t(u)
; NAME/KEY: misc_feature
; OTHER INFORMATION: AAV4 capsid protein VP3 gene
; US-09-532-594B-19

Alignment Scores:
Pred. No.:      8,17e-154      Length:      1617
Score:          1690.50      Matches:      325
Percent Similarity: 71.30%      Conservative: 65
Best Local Similarity: 59.41%      Mismatches:  132
Query Match:      42.38%      Indels:      25
DB:              4          Gaps:        9

```

US-09-807-802A-13 (1-736) x US-09-532-594B-19 (1-1617)

| | | | |
|----|-----|---|-----|
| Qy | 204 | AlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGlyAsn | 223 |
| Db | 7 | GCAGCAGCTGGCGAGCTGCAGTC---GAGGGGGACAAGGTGCGGATGGAGTGGTAAT | 63 |
| Qy | 224 | AlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThrSer | 243 |
| Db | 64 | GCCTCGGGTATTGGCATTCGGATTCCACCTGGTCTGAGGGCCACGTCACGACCACCCAGC | 123 |
| Qy | 244 | ThrArgThrTrpAlaLeuProThrTyTrAsnAsnHisLeuTyTrValGlyGlnIleSerSerAla | 263 |
| Db | 124 | ACCGAACCCTGGGTCTTGCCACCCTCAACCAACCCACCTNTACAAGCGACTCGGAGAG--- | 180 |
| Qy | 264 | SerThrGlyAlaSerAsnAspAsnHisTyTrPheGlyTyTrSerThrProTrpGlyTyTrPhe | 283 |
| Db | 181 | -----AGCCTGCAGTCCACACCTACACCGAATCTCCACCCCTCGGGGATCTTT | 231 |
| Qy | 284 | AspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsnAsn | 303 |
| Db | 232 | GACTTCAACCGCTTCCACTGCCACTTCTCACCGACTGTCGAGCGAGCATCATCAACAC | 291 |
| Qy | 304 | AsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLysGlu | 323 |
| Db | 292 | AACTGGGGCATGCGACCCCAAGCCATCGGGTCAAAATCTTCAACATCCAGTCAAGGAG | 351 |
| Qy | 324 | ValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGlnVal | 343 |
| Db | 352 | GTCAAGCAGCTGAACGGCGAGACAACGGTGGCTAATAACCTTACCAGCACCGTTTCAGATC | 411 |
| Qy | 344 | PheSerAspSerGluTyTrGlnLeuProTyTrValLeuGlySerAlaHisGlnGlyCysLeu | 363 |
| Db | 412 | TTTCGGAGCTGTCGTACGAACTCCCGTACGTGATGATGCGGGTCAAGAGGGCGAGCTTG | 471 |
| Qy | 364 | ProProPheProAlaAspValPheMetIleProGlnTyTrGlyTyTr-----LeuThr | 380 |
| Db | 472 | CCTCTTTTCCCAACGAGCGTCTTATGGTGGCCCCAGTACGGCTACTGTGGAGTGGTGACC | 531 |
| Qy | 381 | LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyTrCysLeuGluTyTrPhePro | 400 |
| Db | 532 | GGCAACACTTCGGAGCAACAGACTGACAGAAATGCCTTCTACTGCTGGAGTACTTCTCT | 591 |

```
Qy 401 SerGlnMetLeuArgThrGlyAsnAsnProPheSerTyrThrPheGluGluValPro 420
Db 592 TCGCAGATGTCGGGACTGGCAACAACCTTGAATACGATACAGTTTGGAGAGTGCT 651
Qy 421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
Db 652 TTCCACTCGATGTACGGCGACAGCAGCGCTGACCGGCTGATGAACCTCTCATCGAC 711
Qy 441 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsn-----GlnSerGlySerAla 456
Db 712 CAGTACTGTGGGAGCTGCAATCATGACCAACACCGGAACCCCTGAATGCCGAGCTGCC 771
Qy 457 GlnAsnLysAspLeuLeuPheSerArgLysSerProAlaGlyMetSerValGlnProLys 476
Db 772 ACCACCAAC-----TTTACCAGCTGGCGCTACCAACTTTTCCAACTTTAAAAAG 822
Qy 477 AsnTrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsn 496
Db 823 AACTGGCTGCCCGGCTTCAATCAAGCAGCAGCGGCTTCTCAAGACTGCC-----AAT 876
Qy 497 AsnAsnSerAsnPheThrTrpThrGlyAlaSer-----LysTyr----- 509
Db 877 CAATACTACAAGATCCCTGCCACCGGGTCACACAGTCTCATCAATACGAGACCCACAGC 936
Qy 510 AsnLeuAsnGlyArgGluSerIleLeuAsnProGlyThrAlaMetAlaSerHisLysAsp 529
Db 937 ACTCTGGACGGAAGATGGAGTGCTGACCCCGGACTTCCCAATGGCCAGCGTGGACCT 996
Qy 530 AspGluAspLysPhePheProMetSerGlyValMetIlePheGlyLysGluSerAlaGly 549
Db 997 CGGACAGCAAG---TTCAGCAACAGCCAGCTCATCTTTGGCGGGGCTTAAACAGACGGC 1053
Qy 550 AlaSerAsnThrAlaLeuAspAsnValMetIleThrAspGluGluIleLysAlaThr 569
Db 1054 AACACGGCCACCGTACCCGGGACTGTGATCTACCTCTGAGAGAGAGCTGGCGCCACC 1113
Qy 570 AsnProValAlaThrGluArgPheGlyThrValAlaValAsnPheGlnSerSerThr 589
Db 1114 AACGCCACCGATACGACATGTGGGGCAACTACCTGCGCGGTGACAGCAACAGCAAC 1173
Qy 590 AspProAlaThrGlyAspValHisAlaMetGlyAlaLeuProGlyMetValTrpLysAsp 609
Db 1174 CTGCGCAGCTGGACAGACTGACAGCTTGGAGCGGTGGCGTGAATGGTCTGCCAAAC 1233
Qy 610 ArgAspValTyrLeuGlnGlyProIleTrpAlaLysIleProHisThrAspGlyHisPhe 629
Db 1234 AGAGACATTTACTACAGGGTCCCATTTGGGCCAAGATCTCTCATCCGATGGACACTTT 1293
Qy 630 HisProSerProLeuMetGlyGlyPheGlyLeuLysAsnProProGlnIleLeuIle 649
Db 1294 CACCCTCACCCTGATTTGGTGGTGGTGGCTGAAACACCCGCTCTCTCAAAATTTTATC 1353
Qy 650 LysAsnThrProValProAlaAsnProProAlaGluPheSerAlaThrLysPheAlaSer 669
Db 1354 AAGAACACCCCGGTACTCGCAATCCTGCAACGACCTTTCAGCTCTACTCCCGTAAACTCC 1413
Qy 670 PheIleThrGlnTyrSerThrGlyValSerValGluIleGluTrpLysGlnLys 689
Db 1414 TTCAATCTCAGTACAGCACTGGCCAGGTGTGGTGGAGATGACTGGGAATCCAGAG 1473
Qy 690 GluAsnSerLysArgTrpAsnProGluValGlnTyrThrSerAsnTyrAlaLysSerAla 709
Db 1474 GAGCGGTCCAAACGCTGGAACCCCGAGGTCCAGTTTACTTCCCACTACGGACAGCAAAAC 1533
Qy 710 AsnValAspPheThrValAspAsnAsnGlyLeuTyrThrGluProArgProIleGlyThr 729
Db 1534 TCTCTGTTGTGGGCTCCCGATCGCGTGGGAAATACACTGAGCCTTAGGGCTATCGGTACC 1593
Qy 730 ArgTyrLeuThrArgProLeu 736
Db 1594 CGTACCTCACCCACCACTG 1614
```

```
RESULT 15
US-09-438-268-3
; Sequence 3, Application US/09438268
; Patent No. 6491907
; GENERAL INFORMATION:
; APPLICANT: Rabinowitz, Joseph E.
; APPLICANT: Samuleki, Richard J
; APPLICANT: Xiao, Weidong
; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING
; FILE REFERENCE: 5470-186
; CURRENT APPLICATION NUMBER: US/09/438,268
; CURRENT FILING DATE: 1999-11-10
; EARLIER APPLICATION NUMBER: 60/107,840
; EARLIER FILING DATE: 1998-11-10
; EARLIER APPLICATION NUMBER: 60/123,651
; EARLIER FILING DATE: 1999-03-10
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2271
; TYPE: DNA
; ORGANISM: Virus
US-09-438-268-3

Alignment Scores:
Pred. No.: 2,07e-126 Length: 2271
Score: 1410.00 Matches: 312
Percent Similarity: 55.75% Conservative: 95
Best Local Similarity: 42.74% Mismatches: 269
Query Match: 35.35% Indels: 54
DB: 4 Gaps: 16
```

```
US-09-807-802A-13 (1-736) x US-09-438-268-3 (1-2271)

Qy 1 MetAlaAlaAspGlyTyrLeuProAspTrpLeuGluAspAsnLeuSerGluGlyIleArg 20
Db 1 ATGGCTGCCGATGGTATCTTCCAGATTGGCTCGAGACACTCTCTCTGAAGGAATAGA 60
Qy 21 GluTrpTrpAspLeuLysProGlyAlaProLysProLysAlaAsnGlnLysGlnAsp 40
Db 61 CAGTGGTGGAGCTCAAACTGGCCACCACCAAGCCCGCAGAGCGGCATAGGAC 120
Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
Db 121 GACAGCAGGGGTCTTGTCTTCTGGGTACAGTACCTCGAGACCTTCAACGAGCTCGAC 180
Qy 61 LysGlyGluProValAsnAlaAlaAspAlaAlaAlaLeuGluHisAspLysAlaTyrAsp 80
Db 181 AAGGGAGAGCCGGTCAACGAGGAGAGCGCCGCGGCTCGAGCAGCAAGACCTACGAC 240
Qy 81 GlnGlnLysLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
Db 241 CGGCAGCTCCAGCGGAGACACCCGCTACTCTCAAGTACAAACCCAGCGCGGAGTTT 300
Qy 101 GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln 120
Db 301 CAGGAGCGCTTAAAGAAGATAGCTCTTTTGGGGCAACCTCGAGCAGCAGTCTCCAG 360
Qy 121 AlaLysLysArgValLeuGluProLeuValGluGluGluAlaLysThrAlaPro 140
Db 361 GCGAAAAAGAGGGGTCTTGAACCTCTGGGCTGGTGGAGAACCTGTATAGACGGCTCCG 420
Qy 141 GlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSerGlyIleGly 160
Db 421 GGAATAAAGAGCGCGGTAGAGCACTCTCTCTGTGGAGCCAGACTCTCTCTCGGGAACCGA 480
Qy 161 LysThrGlyGlnGlnProAlaLysLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
Db 481 AAGCGGGCCAGCAGCTTCAAGAAAAAGATTGAATTTTGGTTCAGACTGGAGACGACAG 540
Qy 181 SerValProAspProGlnProLeuGlyGluProAlaThrProAlaAlaValGlyPro 200
```

Db 541 TCAGTACCTGACCCCGAGCTCTCGACAGCCACCCAGCAGCCCTCTGGTCTGGAACT 600
Qy 201 ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGly 220
Db 601 AATACGATGACTTCAATTAATCTCGAGAA--GCCAGCAGCTGGTGCAGGAGGGGGGC 657
Qy 221 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIle 240
Db 658 AGTAATCTGTCAAAAGCATGTGGAGTGGGGGGCCACTTTAGTGTAACTCTGTAACT 717
Qy 241 ThrThrSerThrArgThrTrpAlaLeuProThrTrpAsnAsnHisLeuTrpIleGlnIle 260
Db 718 TGTACATTTTCCAGACAGTTTAAATTCATATGACCCAGACGACCATTAATAAGGTGTT 777
Qy 261 SerSerAlaSerThrGlyAlaSerAsnAsp----- 270
Db 778 TCTCCCGCAGCAGTAGCTGCCACAATGCCAGTGGAAAGGAGGCAAAAGTTTGGCACCATC 837
Qy 271 AsnHisTyrPheGlyTyrSerThrProTrpGlyTyrPheAspPheAsnArgPheHisCys 290
Db 838 AGTCCCATTAATGGGATCTCAACCCCATGGAGATATTAGATTTTAAATGCTTTAAATTA 897
Qy 291 HisPheSerProArgAspTrpGlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLys 310
Db 898 TTTTTCACCTTTAGAGTTTCAGCAGCTTAATTTGAAATTAATGGAAGTATAGCTCCTGAT 957
Qy 311 ArgLeuAsnPheLysLeuPheAsnIleGlnValLysGluValThr-----ThrAsnAsp 328
Db 958 GCTTTAACTGTAACCATATCAGAAATGTCTTAAAGGATGTTACAGACAAACTCGAGGG 1017
Qy 329 GlyValThrThrIleAlaAsnAsnLeuThrSerThrValGlnValPheSerAspSerGlu 348
Db 1018 GGGGTA---CAGTTACTGACAGCAGCTACAGCGCCCTATGCTATGTTAGTAGACCATGAA 1074
Qy 349 TyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCysLeuProProPheProAla 368
Db 1075 TACAGTACCATATGTGTGTTAGGCAAGGTCAAGTACTTTAGCCCCAGAACTTCTATT 1134
Qy 369 AspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsnAsn---GlySerGlnAla 387
Db 1135 TGGGTATATTTTCCCTCAATATGCTTACTTAACAGTAGGAGATGTTAAACACACAGGA 1194
Qy 388 ValGly-----ArgSerSerPheTyrCysLeuGluTyr 398
Db 1195 ATTTCTGGAGACAGCAAAATTAAGCAAGTGAAGATCAGCAATTTATGTTTGGACAC 1254
Qy 399 PheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGlu 418
Db 1255 AGTTCTTTTTCAGCTTTTAGGTACAGGAGGTACAGCAACTATGCTTATAAGTTTCTCTCA 1314
Qy 419 ValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeu 438
Db 1315 GTGCCCCCAAGAAATTTAGAGGGCTCGCTCAACACTTTTATGAATGTACAACTCCCTTA 1374
Qy 439 IleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsn 458
Db 1375 -----TACGATCCCGCTTAGGGTTCTCGACACATTAGGAGGTGACCCCAAAATTT 1425
Qy 459 LysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrp 478
Db 1426 AGATCTTTA-----ACACATGAAGACCATGCAATTCAGGCCCAAAACTTC 1470
Qy 479 LeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsn 498
Db 1471 ATGCCAGGGCCACTAGTAACCTCAGTGTCTCAAGAGGGGAGACAGCTCTTAATCTGGA 1530
Qy 499 SerAsnPheThrTrpTrpGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIle 518
Db 1531 GCTGGAAAAGCCTTAACAGGCCCTTAGCACAGGTACCTCTCAAAACACTAGATATCTTTA 1590
Qy 519 AsnProGly---ThrAlaMetAlaSerHisLysAspAspGluAspLysPheProMet 537
Db 1591 CGCCCTGGGCCAGTGTCTCAGCCATACCACTGCGGACACAGATAAATATGTCACAGGA 1650

Qy 538 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 557
Db 1651 ATAAATCCATTTCTCATGTGCAGACCACCTATGGT-----AACGCTGAAGACAAA 1701
Qy 558 Val-----MetIleThrAspGluGluGluIleLysAlaThrAsn 570
Db 1702 GAGTATCAGCAGAGGAGTGGTAGATTTCCTCAATGAAAAAGAACAGCTAAACACAGTTACAG 1761
Qy 571 ProValAlaThrGluArgPheGlyThrValAlaValAsnPheGlnSerSerSerThrAsp 590
Db 1762 GGTTTAAACATGCACACTAC-----TTTCCCAATAAAGGAACCCAG 1803
Qy 591 ProAlaThrGlyAspValHisAlaMetGlyAlaLeuProGlyMetValTrpGlnAspArg 610
Db 1804 CAATATACAGATCAAAATTGAG---CGCCCCCTAATGGTGGTCTCTGTATGGACACAGAGA 1860
Qy 611 AspValTyrLeuGlnGlyProIleTrpAlaLysIleProHisThrAspGlyHisPheHis 630
Db 1861 GCCCTTCACTATGAAGCCAGCTGTGGAGTAAATTCCTCAATTTAGATGACAGTTTATAA 1920
Qy 631 ProSer---ProLeuMetGlyGlyPheGlyLeuLysAsnProProGlnIleLeuIle 649
Db 1921 ACTCAGTTTGCAGCCTTTAGGAGGATGGGTTTGCATCAGCCACCTCTCTCAATATTTTA 1980
Qy 650 LysAsnThrProValProAlaAsnProProAlaGluPheSerAlaThrLysPheAlaSer 669
Db 1981 AAA-----ATATTACCACAAAGTGGGCCCAATTGGAGGTATTAAATCAATGGGAATTACT 2034
Qy 670 PheIleThrGlnTyrSerThrGlyGlnValSerValGluIleGluTrpGluLeu---Gln 688
Db 2035 ACCTTAGTTCAAGTATCGCTGGGAATTTATGACAGTAATATGACATTTAAATTTGGGGCCC 2094
Qy 689 LysGluAsnSerLysArgTrpAsnProGlu 698
Db 2095 CGTAAAGCTACGGGACGTTGGAAATCTCTCAA 2124

Search completed: February 15, 2004, 01:34:58
Job time : 286.687 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 14, 2004, 16:55:45 ; Search time 684.807 Seconds
(without alignment)
3959.005 Million cell updates/sec

Title: US-09-807-802a-13

Perfect score: 3989

Sequence: 1 MAADGYPDMLNEDNLSEGR.....NNGLYTEPRPTGTYRLTRPL 736

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPO-spool/US09807802/runat_11022004_175610_15987/app_query.fasta_1.2389
-DB=Published Applications NA -QFWT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=BLOSUM62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09807802 @CNC 1 1 1100 @runat_11022004_175610_15987
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV-TIMEOUT=120 -WARN-TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubna/US09_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubna/US10_NEW_PUB.seq.*
17: /cgn2_6/ptodata/1/pubna/US60_NEW_PUB.seq.*
18: /cgn2_6/ptodata/1/pubna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
|------------|-------|-------------|--------|----|-------------|

| | | | | | | |
|----|--------|-------|------|----|-------------------|-------------------|
| 1 | 3989 | 100.0 | 4718 | 13 | US-10-291-583-6 | Sequence 6, Appl |
| 2 | 3984 | 99.9 | 7447 | 13 | US-10-216-870-11 | Sequence 11, Appl |
| 3 | 3494 | 87.6 | 4726 | 13 | US-10-291-583-8 | Sequence 8, Appl |
| 4 | 3473 | 87.1 | 3122 | 13 | US-10-291-583-45 | Sequence 45, Appl |
| 5 | 3464 | 86.8 | 3117 | 13 | US-10-291-583-43 | Sequence 43, Appl |
| 6 | 3460 | 86.7 | 3121 | 13 | US-10-291-583-44 | Sequence 44, Appl |
| 7 | 3457 | 86.7 | 3122 | 13 | US-10-291-583-42 | Sequence 42, Appl |
| 8 | 3453 | 86.6 | 3128 | 13 | US-10-291-583-27 | Sequence 27, Appl |
| 9 | 3451.5 | 86.5 | 3075 | 13 | US-10-291-583-26 | Sequence 26, Appl |
| 10 | 3448 | 86.4 | 3129 | 13 | US-10-291-583-59 | Sequence 59, Appl |
| 11 | 3446.5 | 86.4 | 3123 | 13 | US-10-291-583-25 | Sequence 25, Appl |
| 12 | 3445 | 86.4 | 3123 | 13 | US-10-291-583-41 | Sequence 41, Appl |
| 13 | 3442.5 | 86.3 | 4721 | 13 | US-10-291-583-1 | Sequence 1, Appl |
| 14 | 3441 | 86.3 | 3128 | 13 | US-10-291-583-47 | Sequence 47, Appl |
| 15 | 3440 | 86.2 | 3128 | 13 | US-10-291-583-28 | Sequence 28, Appl |
| 16 | 3439 | 86.2 | 8179 | 15 | US-10-205-942-5 | Sequence 5, Appl |
| 17 | 3438 | 86.2 | 3084 | 13 | US-10-291-583-39 | Sequence 39, Appl |
| 18 | 3434 | 86.1 | 3098 | 13 | US-10-291-583-37 | Sequence 37, Appl |
| 19 | 3432 | 86.0 | 3197 | 13 | US-10-291-583-29 | Sequence 29, Appl |
| 20 | 3429 | 86.0 | 3121 | 13 | US-10-291-583-11 | Sequence 11, Appl |
| 21 | 3424.5 | 85.8 | 3127 | 13 | US-10-291-583-15 | Sequence 15, Appl |
| 22 | 3423.5 | 85.8 | 2205 | 13 | US-10-291-583-120 | Sequence 120, App |
| 23 | 3423.5 | 85.8 | 4385 | 13 | US-10-291-583-5 | Sequence 5, Appl |
| 24 | 3422.5 | 85.8 | 3095 | 13 | US-10-291-583-24 | Sequence 24, Appl |
| 25 | 3422 | 85.8 | 3121 | 13 | US-10-291-583-13 | Sequence 13, Appl |
| 26 | 3420.5 | 85.7 | 3113 | 13 | US-10-291-583-31 | Sequence 31, Appl |
| 27 | 3420 | 85.7 | 3098 | 13 | US-10-291-583-9 | Sequence 9, Appl |
| 28 | 3417.5 | 85.7 | 3095 | 13 | US-10-291-583-23 | Sequence 23, Appl |
| 29 | 3415.5 | 85.6 | 3122 | 13 | US-10-291-583-56 | Sequence 56, Appl |
| 30 | 3415 | 85.6 | 2489 | 13 | US-10-291-583-35 | Sequence 35, Appl |
| 31 | 3414.5 | 85.6 | 3094 | 13 | US-10-291-583-22 | Sequence 22, Appl |
| 32 | 3414.5 | 85.6 | 3106 | 13 | US-10-291-583-34 | Sequence 34, Appl |
| 33 | 3413.5 | 85.6 | 3123 | 13 | US-10-291-583-54 | Sequence 54, Appl |
| 34 | 3413.5 | 85.6 | 3276 | 13 | US-10-291-583-38 | Sequence 38, Appl |
| 35 | 3413 | 85.6 | 2495 | 13 | US-10-291-583-36 | Sequence 36, Appl |
| 36 | 3411.5 | 85.5 | 3113 | 13 | US-10-291-583-55 | Sequence 55, Appl |
| 37 | 3408.5 | 85.4 | 2504 | 13 | US-10-291-583-33 | Sequence 33, Appl |
| 38 | 3405 | 85.4 | 4393 | 12 | US-10-423-704A-1 | Sequence 1, Appl |
| 39 | 3405 | 85.4 | 4393 | 13 | US-10-291-583-4 | Sequence 4, Appl |
| 40 | 3404 | 85.3 | 3098 | 13 | US-10-291-583-10 | Sequence 10, Appl |
| 41 | 3402.5 | 85.3 | 4679 | 9 | US-09-804-898-1 | Sequence 1, Appl |
| 42 | 3402.5 | 85.3 | 4679 | 9 | US-09-945-681-10 | Sequence 10, Appl |
| 43 | 3402.5 | 85.3 | 4679 | 13 | US-10-136-819-6 | Sequence 6, Appl |
| 44 | 3402.5 | 85.3 | 4679 | 14 | US-10-038-972A-12 | Sequence 12, Appl |
| 45 | 3402 | 85.3 | 3106 | 13 | US-10-291-583-18 | Sequence 18, Appl |

ALIGNMENTS

RESULT 1

US-10-291-583-6
; Sequence 6, Application US/10291583
; Publication No. US2003013872A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alvira, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; TITLE OF INVENTION: Sequences and Isolating No. US2003013872A1
; FILE REFERENCE: US-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; PRIOR FILING DATE: 2002-11-12
; PRIOR FILING DATE: 2001-11-13
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6

```

; LENGTH: 4718.
; TYPE: DNA
; ORGANISM: adeno-associated virus serotype 1
US-10-291-583-6

Alignment Scores:
Pred. No.: 0 Length: 4718
Score: 3989.00 Matches: 736
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 13 Gaps: 0

US-09-807-802A-13 (1-736) x US-10-291-583-6 (1-4718)

QY 1 MetAlaAlaAspGlyTyrLeuProAspTrpLeuGluAspAsnLeuSerGluGlyIleArg 20
DB 2223 ATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGGACAACCTCTCTGAGGGCAATTCGC 2282

QY 21 GluTrpTrpAspLeuLysProGlyAlaProLysProLysAlaAsnGlnLysGlnAsp 40
DB 2283 GAGTGGTGGGACTTGAACCTCGAGCCCGAAGCCCAAGCCCAACAGCAGCAAGCAGGAC 2342

QY 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
DB 2343 GACGGCGGGGCTGTGGTCTTCTGGCTACAAGTACCTCGGACCTTCAACGGACTCGAC 2402

QY 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
DB 2403 AAGGGGAGCCGCTCAACCGCGGAGCGCAGCGGCCCTCGAGCAGCAAGGSCCTACGAC 2462

QY 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
DB 2463 CAGCAGCTCAAGCGGGTGACAATCGGTACCTCGGTATAACACCGCGCAGCCGAGTTT 2522

QY 101 GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyValArgAlaValPheGln 120
DB 2523 CAGGAGCGTCTGCAGAGATACGCTTTTGGGGGCAACTCGGGGAGCAGCTCTTCAG 2582

QY 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaPro 140
DB 2583 GCCAAGAAGCGGTTCTCGAACCCTCTCGGCTCTGGTTGAGGAAGCGCTAAGACGGCTCT 2642

QY 141 GlyLysLysArgProValGlnSerProGlnGluProAspSerSerSerGlyIleGly 160
DB 2643 GGAAGAAAGACGTCCGGTAGCAGTCCGACCAAGAGCCAGACTCTCTCTCGGGCATTCGGC 2702

QY 161 LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
DB 2703 AAGACAGGCGCAGCGCCCTAAAGAGACTCAATTTGGTCAAGCTGGCGACTCAGAG 2762

QY 181 SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro 200
DB 2763 TCAGTCCCCGATCCACAACCTCTCGGAGAACCTCCAGCAACCCCGCTCTGTGGGACCT 2822

QY 201 ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnGluGlyAlaAspGly 220
DB 2823 ACTACAATGGCTTCAGCGGGTGGCGCACCATAATGGCAGACAATAACGAGCGCCGACGGA 2882

QY 221 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIle 240
DB 2883 GTGGGTAAATGGCTCAGGAATTTGGCATTCGGATTCCACATGGCTGGCGCAGAGTCATC 2942

QY 241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIle 260
DB 2943 ACCACGACGCCGACCTGGGCGCTTGGCCACCTACAATAACCCACCTCTACAAGCAATC 3002

QY 261 SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrp 280
DB 3003 TCCAGTGGCTTCAACGGGGGCGCAGCAACGACCACTACTTGGCTTACGACGCCCTCTGG 3062

QY 281 GlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu 300

```

```

DB 3063 GGGTATTTGATTTCAACAGATTCCACTGCCACTTTTCCACCGTGAAGTGGCAGGACTC 3122
QY 301 IleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGln 320
DB 3123 ATCAACAACAATTTGGGATTTCCGGCCCAAGAGACTCAACTTCAAACTCTTCAACATCCAA 3182
QY 321 ValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnLeuLeuSerThr 340
DB 3183 GTCAAGAGGTCACGACGAATGATGGCGTCACACCACTCGCTAATAACCTTTACCGACAG 3242
QY 341 ValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGln 360
DB 3243 GTTCAAGTCTTCTCGGACTCGAGTACAGCTTCCGTACGTCTCGGTCTCGGCACCCAG 3302
QY 361 GlyCysLeuProPheProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr 380
DB 3303 GGCCTGCCTCCCTCGTTCGGCGGAGCGTGTTCATGATTCGGCAATACGGCTACCTGACG 3362
QY 381 LeuAsnAsnGlySerGlnAlaValGlyArgSerPheTyrCysLeuGluTyrPhePro 400
DB 3363 CTCACAATGGCAGCCCAAGCGGTGGAGCTTCACTCTTTACTCGCTGGAAATATTTCCCT 3422
QY 401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 420
DB 3423 TCTCAGATGCTGAGACGGGCAACAACCTTTACCTTCAGCTACACCTTTGAGGAAGTGCCT 3482
QY 421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
DB 3483 TTCCACAGCAGTACCGGCGCACAGCCAGCGCTGAGCGGCTGATGAATTCCTCTCATCGAC 3542
QY 441 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAsp 460
DB 3543 CAATACCTGTATTAACCTGAAACAGAACTCAAAATCAGTCCGGAGTGGCCCAAAACAGGAC 3602
QY 461 LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuPro 480
DB 3603 TTGCTGTTTAGCCGTGGGTCTCCAGCTGGCATGCTGTTCAGCCCAAAACCTGGCTACCT 3662
QY 481 GlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsn 500
DB 3663 GGACCTCTGTATTCGGCAGCAGCGGCTTCTTAAACAAACAAACAGACAAACACACAGCAAT 3722
QY 501 PheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnPro 520
DB 3723 TTTACCTGGACTGTGTCTTCAAAATATAACCTCAATGGCGCTGAATCCATCATCAACCT 3782
QY 521 GlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyVal 540
DB 3783 GGCACCTGCTATGGGCTCACACAAAGACGACGAAAGACAAGTCTCTTCCCATGAGCGGTGTC 3842
QY 541 MetIlePheGlyLysGluSerAlaGlyLysSerAsnThrAlaLeuAspAsnValMetIle 560
DB 3843 ATGATTTTGGAAAGAGAGCGCGGAGCTTCAAAACACTGTCATTTGGACAAATGTCTATGTT 3902
QY 561 ThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrVal 580
DB 3903 ACAGCAGAGAGAGAAATTAAGCCCACTAACCTGTGGCCACCGAAAGATTTGGGACCGTG 3962
QY 581 AlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGly 600
DB 3963 GCAGTCAATTTCCAGAGCAGCAGCAGACACCTTCGACCGGAGATGTGCATGTATGGGA 4022
QY 601 AlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAla 620
DB 4023 GCAATTTACCTGGCATGTGTGGCAAGATAGACAGCTGTAACCTGCGAGGTCCTCCATTTGGGCC 4082
QY 621 LysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeu 640
DB 4083 AAAATTCCTCACACAGATGGACACTTTTCAACCGTCTCTCTTATGGCGGCTTTGGACTC 4142
QY 641 LysAsnProProGlnIleLeuLysAsnThrProValProAlaAsnProAla 660
DB 4143 AAGAACCGGCTCTCTCAGATCTCATCAAAACACGCGCTGTCTCTGCGGAATCTCTCGGGC 4202

```

Qy 661 GluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSer 680
 Db 4203 GAGTTTTCAGTACAAAGTTTGGCTTCTCATCACCACCACTCATCCAGGACAGTGTGAGT 4262
 Qy 681 ValGluIleGluTrpGluLeuGlnLysGluAenSerLysArgTrpAenProGluValGln 700
 Db 4263 GTGGAAATTGAATGGGAGCTGCAGAAAGAAACAGCAAGCGCTGGAATCCGGAAGTGCAG 4322
 Qy 701 TyrThrSerAenTyrAlaLysSerAlaAenValAenPheThrValAenAenGlyLeu 720
 Db 4323 TACATCCCAATTATGCAAAATCTGCCAACGTTGATTTTACTGTGGACAAATGGAGCTT 4382
 Qy 721 TyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
 Db 4383 TATACTGAGCTCGCCCATTTGGCACCCTGTACCTTACCCGCTGCCCTG 4430
 RESULT 2
 US-10-216-870-11
 ; Sequence 11, Application US/10216870
 ; Publication No. US20030148506A1
 ; GENERAL INFORMATION:
 ; APPLICANT: KOTIN, ROBERT M
 ; APPLICANT: URABE, MASASHI
 ; TITLE OF INVENTION: PRODUCTION OF ADENO-ASSOCIATED VIRUS IN INSECT CELLS
 ; FILE REFERENCE: 402133
 ; CURRENT APPLICATION NUMBER: US/10/216.870
 ; CURRENT FILING DATE: 2002-08-13
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 11
 ; LENGTH: 7447
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Synthetic
 US-10-216-870-11
 Alignment Scores:
 Pred. No.: 0 Length: 7447
 Score: 3984.00 Matches: 735
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.87% Indels: 0
 DB: 13 Gaps: 0
 US-09-807-802A-13 (1-736) x US-10-216-870-11 (1-7447)
 Qy 2 AlaAlaAspGlyTyrLeuProAspTrpLeuGluAenLeuSerGluGlyIleArgGlu 21
 Db 4624 GCTGCCGACGGTTATCTACCCGATTTGGCTCGAGGACCACTCTCTGAGGGCATTCGCGAG 4683
 Qy 22 TrpTrpAspLeuLysProGlyAlaProLysProLysAlaAenGlnLysGlnAsp 41
 Db 4684 TGGTGGGACTTGAACCTGGAGCCCGAAGCCCAAGCCAAACGACAAAGACGACGAC 4743
 Qy 42 GlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAenGlyLeuAspLys 61
 Db 4744 GGCCGGGGTCTGGTGTCTTCTCGGCTACAAAGTACCTCGGACCCCTTCAACGGACTCGACAAG 4803
 Qy 62 GlyGluProValAenAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAspGln 81
 Db 4804 GGGGAGCCCGTCAACGGCGGCGAGCGAGCGCCCTCGAGCAGCAGCAAGGCCCTACGACCAG 4863
 Qy 82 GlnLeuLysAlaGlyAspAenProTyrLeuArgTyrAenHisAlaAenAlaGluPheGln 101
 Db 4864 CAGCTCAAGGGGTGACAAATCCGTACCTCGGCTATACCAACGCGCGAGCTTTCAG 4923
 Qy 102 GluArgLeuGlnGluAspThrSerPheGlyGlyAenLeuGlyArgAlaValPheGlnAla 121
 Db 4924 GAGCGTCTGCAGAAAGATACGTCTTTTGGGGGCAACCTTCGGGCGAGCAGTCTTCCAGGCC 4983

Qy 122 LysLysArgValLeuGluProLeuGlyLeuValGluGluAlaLysThrAlaProGly 141
 Db 4984 AAGAAGCGGGTCTCGAACCTCTCGGTCTGGTTGAGAAAGCGCTAAGACGGCTCTCTGGA 5043
 Qy 142 LysLysArgProValGluGlnSerProGlnGluProAspSerSerSerGlyIleGlyLys 161
 Db 5044 AAGAAACGTCGGGTAGAGCAGTCGCCACAAGACGACAGCTCTCTCTCGGGCATCGGCAAG 5103
 Qy 162 ThrGlyGlnGlnProAlaLysLysArgLeuAenPheGlyGlnThrGlyAspSerGluSer 181
 Db 5104 ACAGGCGCAGAGCCCGCTAAAGAGAGACTCAATTTTGGTCAGACTGGCGACTCAGAGTCA 5163
 Qy 182 ValProAspProGlnProLeuGlyGluProProAlaThrProAlaLysValGlyProThr 201
 Db 5164 GTCCCGATCCACAACCTCTCGGAGAACCTCCAGCAACCCCGCTGCTGTGGGACCTACT 5223
 Qy 202 ThrMetAlaSerGlyGlyAlaProMetAlaAspAenGluGlyAlaAenGlyVal 221
 Db 5224 ACAATGGCTTCAGGCGGTGGCGCACCAATGGCGACAAATAACGAAGCGCGCAGGAGTG 5283
 Qy 222 GlyAenAlaSerGlyAenTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThr 241
 Db 5284 GGTAATGCCCTCAGAAATTTGGCATTTCCGATTTCCACATGGCTGGCGACAGAGTCATCACC 5343
 Qy 242 ThrSerThrArgThrTrpAlaLeuProThrTyrAenAenHisLeuTyrLysGlnIleSer 261
 Db 5344 ACCAGCACCGCACCTTGGCGCTTGGCCACCTACAATAACCCACTCTACAAGCAAACTCC 5403
 Qy 262 SerAlaSerThrGlyAlaSerAenAenAenHisTyrPheGlyTyrSerThrProTrpGly 281
 Db 5404 AGTGCTTCAACGGGGCGCACCAACCAACCACTACTTCCGCTACAGCACCCCTCGGGG 5463
 Qy 282 TyrPheAspPheAenArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIle 301
 Db 5464 TATTTGATTTCAACAGATTTCCACTGCCACTTTTCCACCAGTGCCTGGCGACGACTCATC 5523
 Qy 302 AenAenAenTrpGlyPheArgProLysArgLeuAenPheLysLeuPheAenIleGlnVal 321
 Db 5524 AACAAATTTGGGATTCGGGCCCAAGAGACTCAACTTCAAACTCTTCAACATCAAGTC 5583
 Qy 322 LysGluValThrThrAenAspGlyValThrThrIleAenAenLeuThrSerThrVal 341
 Db 5584 AAGGAGTTCACGACGAATGATGGCGTCACAAACCATCGCTAATAACCTTACGACACGGTT 5643
 Qy 342 GlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGly 361
 Db 5644 CAAGTCTTCTCGGACTCGGAGTACAGCTTCGTAAGTCTCGGCTCGGCGACCGGCGC 5703
 Qy 362 CysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeu 381
 Db 5704 TGCTCTCCCTCCGTTCCCGCGGACGTGTTCATGATTCGCAATACGCTACCTGACGCTC 5763
 Qy 382 AenAenGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSer 401
 Db 5764 AACAAATGGGACCAAGCGGTGGGACCTTCATCTCTTACTGCTGGAATATTTCCCTTCT 5823
 Qy 402 GlnMetLeuArgThrGlyAenAenPheThrPheSerTyrThrPheGluGluValProPhe 421
 Db 5824 CAGATGCTGAGAACGGGCAACAACTTACCTTCAGCTACACCTTTGAGGAAGTGCCTTTC 5883
 Qy 422 HisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAenProLeuIleAenGln 441
 Db 5884 CACAGCAGCTACCGCACACGACGAGCTGGACCGCTGATGAATCTCTCTCATCGACCAA 5943
 Qy 442 TyrLeuTyrTyrLeuAenArgThrGlnAenGlnSerGlySerAlaGlnAenLysAenLeu 461
 Db 5944 TACTGTATTACTTGAAACAGAACTCAAAATCAGTCCGGAAGTCCCAAAACAGAGACTTG 6003
 Qy 462 LeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAenTrpLeuProGly 481
 Db 6004 CTGTTTAGCGGTGGTCTCCAGCTGGCATGTCTGTTTCAGCCCAAAACCTGCTACCTGGA 6063
 Qy 482 ProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAenAenAenSerAsnPhe 501

```
Db 6064 CCCTGTATTGCGCAGCGCGCTTTCTAAACAAACACAGACACACACGCAATTTT 6123
Qy 502 ThrTrpThrGlyValSerLysTyrAenLeuAenGlyArgGluSerLysLeileAenProGly 521
Db 6124 ACCTGGACTGGTGTCTCAAAATATAACCTCAATGGCGTGAATCCATCATCAACCTGGC 6183
Qy 522 ThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMet 541
Db 6184 ACTGCTATGGCTCCACAAACACGACGAAAGCAAGTTCTTTCCCATGAGCGGTGTATG 6243
Qy 542 IlePheGlyLysGluSerAlaGlyAlaSerAenThrAlaLeuAenAspValMetIleThr 561
Db 6244 ATTTTTGGAAAGAGAGCGCCGAGCTTCAACACTGCATTGGCAATGTCTATGATTACA 6303
Qy 562 AspGluGluLysAlaThrAenProValAlaThrGluArgPheGlyThrValAla 581
Db 6304 GACGAAGAGGAAATTAAGCCACTAACCTGTGGCCACCGAAAGATTTGGGACCGTGCA 6363
Qy 582 ValAenPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyVala 601
Db 6364 GTCNAATTTCCAGACGACGACGACGACCTGCGACCGGAGATGTGCATGCTATGGGACA 6423
Qy 602 LeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLys 621
Db 6424 TTACCTGGCATGTGTGGCAAGATAGACGCTGACCTGACGGTCCCATTTGGGCCCAA 6483
Qy 622 IleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLys 641
Db 6484 ATTCTCTCACACAGATGGACACTTTACCCGCTCTCTCTTATGGCGGCTTTGGACTCAAG 6543
Qy 642 AenProProGlnIleLeuLysAenThrProValProAlaAenProProAlaGlu 661
Db 6544 AACCCGCCCTCCAGATCTCATCAAAACACGCTGTCTTCCGCAATCTCCGCGCGAG 6603
Qy 662 PheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerVal 681
Db 6604 TTTTCAGCTACAAAGTTTGCTTCATTCATCCCACTACTCCACAGGACAAAGTAGTG 6663
Qy 682 GluIleGluTrpGluLeuGlnLysGluAenSerLysArgTrpAenProGluValGlnTyr 701
Db 6664 GAAATTTGAATGGAGCTGCAGAAAGAAAAACAGCAAGCGCTGGAATCCCGAAGTGCAGTAC 6723
Qy 702 ThrSerAenTyrAlaLysSerAlaAenValAspPheThrValAenAenAenGlyLeuTyr 721
Db 6724 ACATCCAAATATGCAAAATCTGCCAACGTTGATTTTACTGTGGCAACAAATGGACTTTAT 6783
Qy 722 ThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 6784 ACTGAGCTCGCCCATTTGCACCCGCTTACCTTTACCCGCTCCCTG 6828

RESULT 3
US-10-291-583-8
; Sequence 8, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alvira, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; FILE REFERENCE: UPN-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
```

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 4726
; TYPE: DNA
; ORGANISM: adeno-associated virus serotype 3
US-10-291-583-8

Alignment Scores:
Pred. No.: 0 Length: 4726
Score: 3494.00 Matches: 637
Percent Similarity: 91.72% Conservative: 39
Best Local Similarity: 86.43% Mismatches: 59
Query Match: 87.59% Indels: 2
Db: 13 Gaps: 2
```

US-09-807-802A-13 (1-736) x US-10-291-583-8 (1-4726)

```
Qy 1 MetAlaAlaAspGlyTyrLeuProAspTrpLeuGluAenLeuSerGluGlyIleArg 20
Db 2209 ATGGCTCTGACGGTTATCTTCCAGATTGGCTCGAGGACAACCTTTCTGAAGGCATTCGT 2268
Qy 21 GluTrpTrpAspLeuLysProGlyAlaProLysProllysAlaAenGlnGlnLysGlnAsp 40
Db 2269 GAGTGGTGGGCTCTGAAACCTGGAGTCCCTCAACCCAAAGCGAACCAACCAACCCAGGAC 2328
Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAenGlyLeuAsp 60
Db 2329 AACCGTGGGGTCTTGCTTCCGGTTACAAATACCTCGGACCCCGGTAAACGACTCGAC 2388
Qy 61 LysGlyGluProValAenAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
Db 2389 AAAGAGAGCGCGTCAACAGGCGGACGCGGACGCTCGAAACACGACAAAGCTTACGAC 2448
Qy 81 GlnGlnLeuLysAlaGlyAspAenProTyrLeuArgTyrAenHisAlaAenAlaGluPhe 100
Db 2449 CAGCAGCTCAAGCGCGGTGACCAACCGTACCTCAAGTACCAACACCGCGCGAGGTTT 2508
Qy 101 GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAenLeuGlyArgAlaValPheGln 120
Db 2509 CAGGAGCGCTTCAAGAGATACCTCTTTGGGGGCAACCTTTGGCAGAGCAGTCTTCCAG 2568
Qy 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaPro 140
Db 2569 GCCAAAAAGAGATCTTTAGCCTCTTGGTCTGGTGAAGAGCAGCTTAAACCGCTCCT 2628
Qy 141 GlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSerGlyIleGly 160
Db 2629 GGAAAGAGGGGGGTGTAGATCAGTCTCTCAGGAACCGGACTCATCATCTGTGTGTGGC 2688
Qy 161 LysThrGlyGlnGlnProAlaLysLysArgLeuAenPheGlyGlnThrGlyAspSerGlu 180
Db 2689 AAATCGGGCAACAGCCTGCCAGAAAAAGACTAAATTTCCGTTCAGCTGGAGACTCAGAG 2748
Qy 181 SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro 200
Db 2749 TCAGTCCCGACACCTCTCAACCTCTCGGAGAACCCAGCAGCCCCCACAAGTTTGGGATCT 2808
Qy 201 ThrThrMetAlaSerGlyGlyValaProMetAlaAspAenAenGluGlyAlaAspGly 220
Db 2809 AATCAATGGCTTACGGCGGTGGCCACCAATATGCGACGACATATACGAGGTGCGCATGGA 2868
Qy 221 ValGlyAenAlaSerGlyAenTrpHisCysAspSerThrTrpLeuGlyAspArgValle 240
Db 2869 GTGGGTAAATCTCTCAGGAAATTTGGCATTTCCCAATGGCTCCGCGGACAGAGTCAATC 2928
Qy 241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAenAenHisLeuTyrLysGlnIle 260
Db 2929 ACCACGACGACGACGACCTGGGCCCTTCCCTTACAAACCAACATCTCTTACAGCAAAATC 2988
Qy 261 SerSerAlaSerThrGlyAlaSerAenAenHisTyrPheGlyTyrSerThrProTrp 280
Db 2989 TCCAGCCCAATCA---GGAGCTTCAAAACGACCACTACTTTGGCTACAGCAGCCCCCTTGG 3045
```


QY 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
DB 1084 CAGCAGCTCAAGCGGGTGACAACTCGTACCTGCGGTATAATCAACGCGGAGCGCGAGTTT 1143
QY 101 GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln 120
DB 1144 CAGGAGCGTCTCAAGAGATACGTCTTTTGGGGGCAACCTCGGGGAGCAGTCTTCAG 1203
QY 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaPro 140
DB 1204 GCCAAGAGCGGTTCTCGAACCCTCTCGGTCTGGTTGAGGAGGCGCTAAGACGCGTCTCT 1263
QY 141 GlyLysLysArgProValGlnGlnSerProGlnGluProAspSerSerSerGlyLeuGly 160
DB 1264 GGAAAGAGAGACCGGTAGCAGTCCGCCAAGAGCCAGACTCTCTCTCGGGGCACTCGC 1323
QY 161 LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
DB 1324 AAGACAGCCAGCAGCGCCCTAAAGAGACTCAATTTTGGTCAGACTGGCGACTCAGAG 1383
QY 181 SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro 200
DB 1384 TCAGTCCCGAGCCCAACACTCTCGGAGAACCTCCAGCAGCCCTCAGGTCTGGGAGCT 1443
QY 201 ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnGluGlyAlaAspGly 220
DB 1444 AATACAAATGGCTTCAGCGGTGGCGCTCCAAATGGCAGACAAATACGAAGCGCCGACGGA 1503
QY 221 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIle 240
DB 1504 GTGGGTAAATCTCGGGAAATTTGGCATTGGATTCCACATGGCTGGGGACAGAGTCATC 1563
QY 241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrIleGlnIle 260
DB 1564 ACCACAGCAGCCCGAACCTTGGGCCCTCGCCACCTCAACAACACCACTCTACAGCAAAATC 1623
QY 261 SerSerAlaSerThrGlyAlaSer---AsnAspAsnHisTyrPheGlyTyrSerThrPro 279
DB 1624 TCCACGGCACCTTCGGGAGGAGCACCACAGCACACACTATTTTGGCTACAGCACCCCC 1683
QY 280 TrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArg 299
DB 1684 TGGGGGTATTTTGACTTCAACAGATTCCACTGTCTCACTTTTCCACAGCTGACGGAACGA 1743
QY 300 LeuIleAsnAsnMetTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIle 319
DB 1744 CTCATCAACAACAATTTGGGGATTTCGGGCCCAAAAGACTCAACTTCAAGCTGTTCACATC 1803
QY 320 GlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSer 339
DB 1804 CAGETCAAGGAAGTCAAGCAGAACGAGGACCAACAGCCATCGCCATTAATCTCACCAGC 1863
QY 340 ThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHis 359
DB 1864 ACCGTGCGAGTCTTTACGGACTCGAGTACCAAGTACCGTACGTAGGATCCGCTCAC 1923
QY 360 GlnGlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeu 379
DB 1924 CAGGGAATGTCTGCTTCGTTCGGGGAGCGTCTTATGGTTCTCAGTACGGCTATTTA 1983
QY 380 ThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhe 399
DB 1984 ACTTTAAACATAGAGACCAAGCCCTGGGAGCGTCTCTCTTACTGTCTGGAGTATTTTC 2043
QY 400 ProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluVal 419
DB 2044 CCATCGCAGATGCTGAGAACCGGCAACACTTTTCAGTTCAGCTACACTTCGAGGACGTG 2103
QY 420 PropheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIle 439
DB 2104 CCTTTCCACAGCAGCTACGCGCACGACGAGCGCTGGACAGCGGTGATGAATCCCCCTCATC 2163

QY 440 AspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLys 459
DB 2164 GACCAGTACCTGTACTACCTGGTCAGAACGCAACAG---ACTGGAATCGGGAGCAGCAG 2220
QY 460 AspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeu 479
DB 2221 ACTCTGGCATTCAGCAAGCGGGTCTTAGCTCAATGCGCAACACGAGGTAGAAATGGGTG 2280
QY 480 ProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSer 499
DB 2281 CCGGAGCTTTGCTACCGGCGACGACGCGCTCTCCACGACCAACCAACAGAACCAACAGC 2340
QY 500 AsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsn 519
DB 2341 AACTTTGCTTGGACGGGAGCTGCCAAGTTTAAGCTGAACGGCGAGACTCTCTAATGAAT 2400
QY 520 ProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGly 539
DB 2401 CCGGCGGTGGCAATGGCTTCCCAAGGATGACGACGACGCGCTCTTCTCCCTTCAGCGGG 2460
QY 540 ValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMet 559
DB 2461 GTCTGTATTTTGGCAAGCAAGGAGCGGGAAACGATGGAGTGGATTACAGCAAGTGTCTG 2520
QY 560 IleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThr 579
DB 2521 ATTACAGATGAGGAAGAAATCAAGCTTACCACCCCGTGGCCACAGAAGATATGGAGCA 2580
QY 580 ValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValIleAlaMet 599
DB 2581 GTGGCCATCAACACCAAGCGGCCAATACGACGCGCAGACCGGACTCGTGCAACACCG 2640
QY 600 GlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrp 619
DB 2641 GGGGTGATTTCCCGGATGGTGTGGCAGAAATAGACAGCTGTACCTGACGGGTCCCATCTGG 2700
QY 620 AlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGly 639
DB 2701 GCCAAATTCCTCACAGGACGGCAACTTTCACCGCTCTCCCTGATGGCGGCTTTGGA 2760
QY 640 LeuLysAsnProProGlnIleLeuLysAsnThrProValProAlaAsnProPro 659
DB 2761 CTGAAGCACCGCGCTCTCAATTTCTCATCAAGAACACACCGGTTCACGCGGACCCGCG 2820
QY 660 AlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnVal 679
DB 2821 CTTACTTCAACAGGCGCAAGCTGAACCTTTTCATCAGCAGTACAGACCCGACAGTCTC 2880
QY 680 SerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluVal 699
DB 2881 AGCGTGGAAATCGAGTGGAGCTGCAGAAAGAAACAGCAACGCTGGAATCCAGAGATT 2940
QY 700 GlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGly 719
DB 2941 CAATACACTTCCAATCTACTACAAATCTACAATGTGGACTTGTCTGTCAACAGGAGGG 3000
QY 720 LeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
DB 3001 GTTTATAGGAGCTCGCCCATTTGGCACCGCTTACCTCACCGCAACCTG 3051

RESULT 5

US-10-291-583-43

; Sequence 43, Application US/10291583

; Publication No. US20030138772A1

; GENERAL INFORMATION:

; APPLICANT: Gao, Guangping

; APPLICANT: Wilson, James M.

; APPLICANT: Alvira, Mauricio

; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus

; FILE REFERENCE: UPN-02735USA

; CURRENT APPLICATION NUMBER: US/10/291,583

; CURRENT FILING DATE: 2002-11-12

;; PRIOR APPLICATION NUMBER: US 60/350,607
;; PRIOR FILING DATE: 2001-11-13
;; PRIOR APPLICATION NUMBER: US 60/341,117
;; PRIOR FILING DATE: 2001-12-17
;; PRIOR APPLICATION NUMBER: US 60/377,066
;; PRIOR FILING DATE: 2002-05-01
;; PRIOR APPLICATION NUMBER: US 60/386,675
;; PRIOR FILING DATE: 2002-06-05
;; NUMBER OF SEQ ID NOS: 120
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 43
;; LENGTH: 3117
;; TYPE: DNA
;; ORGANISM: 43.21
US-10-291-583-43

Alignment Scores:

| | | | |
|------------------------|---------|---------------|------|
| Pred. No.: | 0 | Length: | 3117 |
| Score: | 3464.00 | Matches: | 629 |
| Percent Similarity: | 91.59% | Conservative: | 46 |
| Best Local Similarity: | 85.35% | Mismatches: | 60 |
| Query Match: | 86.84% | Indels: | 2 |
| DB: | 13 | Gaps: | 2 |

US-09-807-802A-13 (1-736) x US-10-291-583-43 (1-3117)

Qy 1 MetAlaAlaAspGlyLeuProAspTyrLeuGluAspAsnLeuSerGluGlyLeuArg 20
Db 839 ATGGCTGCCGATGGTATCTCCAGATTGGCTCGAGGACCACTCTCTGAGGGCATTCGC 898
Qy 21 GluTrpTrpAspLeuLeuProGlyAlaProLysProLysAlaAsnGlnGlnLysGlnAsp 40
Db 899 GAGTGGTGGGACTTGAACCTGGAGCCCGGAAACCCAAAGCCCAACACGAGCAAGCAGC 958
Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
Db 959 GACGCGCGGGTCTGGTCTCTCGCTTACAGTACCTCGGACCTTCAACGGACTCGAC 1018
Qy 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
Db 1019 AAGGGGAGCCGTCNACCGCGGAGCCGAGCGGCCCTCGAGCAGCAGCAAGCCTACGAC 1078
Qy 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
Db 1079 CAGCAGCTCAAAAGCGGTGACAATCCGTACCTCGGTATATATCAACGCGCAGCGCGAGTTT 1138
Qy 101 GlnGluArgLeuGlnGluAspThrPheGlyGlyAsnLeuGlyArgAlaValPheGln 120
Db 1139 CAGGAGCGTCTGCAAGAGATACGTCTTTTGGGGGCAACCTCGGGCGAGCAGTCTTCCAG 1198
Qy 121 AlalyLysArgValLeuGluProLeuGlyLeuValGluGlyAlaLysThrAlaPro 140
Db 1199 GCCAAGAGCGGTTCTCGAACCTCTCGTCTGGTTGAGGAAGCGCTTAAGCGGCTCT 1258
Qy 141 GlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerGlyLysGly 160
Db 1259 GGAAGAAAGAGACCGGTAGAGCAGTCGCCACAAAGACGACAGCTCTCTCGGGCATCGGC 1318
Qy 161 LysThrGlyGlnGlnProAlalyLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
Db 1319 AAGACAGCCAGCAGCCCGCTAAAGAGACTCAATTTTGGTTCAGACTGGGCACTCAGAG 1378
Qy 181 SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro 200
Db 1379 TCAGTCCCGACCCACACCTCTCGAGAACCTCCAGCAGCCCTCAGGCTCGGACCT 1438
Qy 201 ThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnGlnGlyAlaAspGly 220
Db 1439 AATACAATGGCTTTCAGGCGGTGGCGCTCCAAATGGCAGACAATAACGAAGCGCCACGGA 1498
Qy 221 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTroLeuGlyAspArgValIle 240
Db 1499 GTGGGTAAATTCCTCGGGAATTTGGCAATTCGAGATTCACATGCTGGGGACAGAGTCATC 1558

Qy 241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIle 260
Db 1559 ACCACCAAGCACCAGACCTGGGCGCTGCCACCTACCAACAACCACTCTCAACAACAATC 1618
Qy 261 SerSerAlaSerThrGlyAlaSer---AsnAspAsnHisTyrPheGlyTyrSerThrPro 279
Db 1619 TCCAACGGCACCTCGGAGGAGAACCAACAGCAGACCACTATTATTGGCTCAGCACCCTCC 1678
Qy 280 TrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArg 299
Db 1679 TGGGGGTATTTCCTTCAACAGATTCCATCTGTACCTTTTCCACCACTGACTGGCAACGA 1738
Qy 300 LeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIle 319
Db 1739 CTCATCAACAACAATTTGGGATTTCCGGCCCAAGAAGCTCAACTTCAAGCTGTTCAACATC 1798
Qy 320 GlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnLeuThrSer 339
Db 1799 CAGGTCAAGGAGTCAAGCAGAACGAGGACCAACCACTGCCAATATCTCACCAGC 1858
Qy 340 ThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHis 359
Db 1859 ACCGTGGGGTCTTTACGGACTCGAGTACAGTTACCGTACGCTAGTACGATCCGCTCAC 1918
Qy 360 GlnGlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeu 379
Db 1919 CAGGATGTCCTGCTCCGTTCCCGGGGACGCTTTCATGCTTCCTAGTACGGCTATTTA 1978
Qy 380 ThrLeuAsnGlnSerGlnAlaValGlyArgSerSerPheTyrCysLeuGlyTyrPhe 399
Db 1979 ACTTTAAACAATGGAAGCCCAAGCCCTGGGACGTTCTCTCTTACTCTCTGGAGTATTTC 2038
Qy 400 ProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluVal 419
Db 2039 CCATCGCAGATGCTGAGAACCGGCACACACTTTCAGTTTCACTACACTTCGAGGAGCTG 2098
Qy 420 ProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIle 439
Db 2099 CTTTTCACAGCAGTACCGGCACAGCAGCGCTGACAGGCTGATGAATCCCTCTCATC 2158
Qy 440 AspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLys 459
Db 2159 GACCAGTACCTGTACTACTCTGGTCAAGAACGAAACG---ACTGGAACTGGAGGACGCG 2215
Qy 460 AspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeu 479
Db 2216 ACTCTGCATTCAGCCAAAGCGGCTCTAGCTCAATGSCCAACCAAGGCTAGAAATGGGTG 2275
Qy 480 ProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSer 499
Db 2276 CCGGACCTTGTACCGGCAGCAGCGGCTCTCCACGACCAACCAACAGAGCAACAACAGC 2335
Qy 500 AsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsn 519
Db 2336 AACTTTCCTCGGAGCGGAGCTCCCAAGTTTAAAGCTGAACCGCGGAGACTCTCTAATGANT 2395
Qy 520 ProGlyThrAlaMetAlaSerHisLysAspGluAspLysPhePheProMetSerGly 539
Db 2396 CCGGCGGTGGCAATGGCTTCCACAGAGATGACAGCAGCGCTTCTTCCCTTCGAGCGGG 2455
Qy 540 ValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMet 559
Db 2456 GTCTCTGATTTTGGCAAGCAGGAGCGGGAAACGATGGAGTGGATTACAGCAAGTGTCTG 2515
Qy 560 IleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThr 579
Db 2516 ATTACAGATGAGGAAGAAATCAAGGCTACCAACCCCGTGGCCACAGAAATATGGAGCA 2575
Qy 580 ValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValIleAlaMet 599
Db 2576 GTGGCCATCAACAACACGAGCGGCCCAATACGCGGCGGACGCGGACTCGTGCACAACAGC 2635


```
QY 600 GlyAlaLeuProGlyMetValTrpGlnAsePArgAspValTyrLeuGlnGlyProIleTrp 619
Db 2636 GGGGTGATTCGGCGATGGTGGGAGAAATAGAGACGTGTACTCGAGGGTCCCATCTGG 2695
QY 620 AlaValIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGly 639
Db 2696 GCCAAATTCCTCACAGGACGGCAACCTTTACCCCGTCTCCCTCGATGGGCGCTTGGGA 2755
QY 640 LeuYsAenProProGlnIleLeuIleYsAenThrProValProAlaAenProPro 659
Db 2756 CTGAAGCACCGCCCTCCTCAAAATCTCATCAAGAACACACACCGGTTCCAGCGGACCGCG 2815
QY 660 AlaGluPheSerAlaThrYsPheAlaSerPheIleThrGlnTyrSerThrGlyGlnVal 679
Db 2816 CTTACTCTCAACACGCGCCAAAGCTGAATCTTTTCATCAGCAGTACAGACCGGACAGGTC 2875
QY 680 SerValGluIleGluTrpGluLeuGlnYsGluAenSerLysArgTrpAenProGluVal 699
Db 2876 AGCGTGGAAATCAGTGGGAGCTGCAGAAAGAAACAGCAACGCTGGAAATCCAGAGATT 2935
QY 700 GlnTyrThrSerAsnTyrAlaYsSerAlaAenValAspPheThrValAspAenGly 719
Db 2936 CAATACACTTCCAACTACTACAAATCTACAAATGTGGACTTTGTGTCAACACGGAAGGA 2995
QY 720 LeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 2996 GTTTATAGGAGCCTCGCCCATTTGGCACCCGTTACCTCACCCTGCAACCTG 3046

RESULT 6
US-10-291-583-44
; Sequence 44, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alvira, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; TITLE OF INVENTION: Sequences and Isolating No. US20030138772A1el Sequences Identifi
; FILE REFERENCE: UPN-0273505A
; CURRENT APPLICATION NUMBER: US/10/291,583
; PRIOR FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 3121
; TYPE: DNA
; ORGANISM: 43.23
US-10-291-583-44

Alignment Scores:
Pred. No.: 0 Length: 3121
Score: 3460.00 Matches: 629
Percent Similarity: 91.32% Conservatives: 44
Best Local Similarity: 85.35% Mismatches: 62
Query Match: 86.74% Indels: 2
DB: 13 Gaps: 2

US-09-807-802A-13 (1-736) x US-10-291-583-44 (1-3121)
QY 1 MetAlaAsePglyTyrLeuProAsePTrpLeuGluAsePAsnLeuSerGluGlyIleArg 20
Db 843 ATGGCTGGCGATGGTATCTTCAGATTGGCTCGAGGACAACTCTCTGAGGCGCATTCG 902
QY 21 GluTrpTrpAspLeuYsProGlyAlaProLysProLysAlaAenGlnGlnYsGlnAseP 40
```

```
Qy 400 ProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluVal 419
Db 2043 CCATCGCAGATGCCGAGACCGGCACAACTTTTCAGTTCAGTACACCTTCGAGGAGGTG 2102
Qy 420 ProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIle 439
Db 2103 CCTTTCACAGCAGCTACGCGCACAGCCAGAGCCTGGACAGGCTGATGAATCCCTCATC 2162
Qy 440 AspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLys 459
Db 2163 GACCAGTACCTGTACTACCTGGTCAGAACGCAACG---ACTGGAACCTGGAGGACCGCAG 2219
Qy 460 AspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeu 479
Db 2220 ACTCTGGCATTCAGCAACGCGGTCTTAGCTCAATGGCCACACCGCTAGAGATTGGGTG 2279
Qy 480 ProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSer 499
Db 2280 CCGGACCTTGCTACCGCAGCAGCGCTCTCCACGACCAACCAACCAACCAACCAACAGC 2339
Qy 500 AsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleAsn 519
Db 2340 AACTTTGCTGGACGGGAGCTGCCAAGTTTAAGCTGAACGCGCCGAGACTCTCTAATGAAT 2399
Qy 520 ProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGly 539
Db 2400 CCGGGCTGGCAATGGCTTCCACADAGATGACGACGACCGCTTCTTCCCTTCAGCGGG 2459
Qy 540 ValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMet 559
Db 2460 GTCTGTGATTTTGGCAAGAGGACCGGCAACGATGAGTGGATTACAGCAAGTGTCTG 2519
Qy 560 IleThrAspGluGluGluIleValAlaThrAsnProValAlaThrGluArgPheGlyThr 579
Db 2520 ATTACAGATAGGAAGAAATCAAGCTACCAACCCCGTGGCCACAGAAGATATGGAGCA 2579
Qy 580 ValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMet 599
Db 2580 GTGGCCATCAACACACGCGCCCAATACGAGCGCGCACACCGACTCGTGCACACACAG 2639
Qy 600 GlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrp 619
Db 2640 GGGGTGATTCGCGCATGGTGTGGCAGAAATAGACAGTGTACTCTGCAGGGTCCCATCTGG 2699
Qy 620 AlaValIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGly 639
Db 2700 GCCAAATTCCTCACACGCGCGCAACTTTCACCCGCTCTCCCTGATGGCGGCTTTTGA 2759
Qy 640 LeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsnProPro 659
Db 2760 CTGAGACACCGCCCTCTCAATTCATCAAGAACACACCGGTTCCAGCGGACCGCGCG 2819
Qy 660 AlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnVal 679
Db 2820 CTTACTCTCAACAGCGCCCAAGCTGAACCTTTTCATCAGCGAGTACAGCACCGGACAGT 2879
Qy 680 SerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluVal 699
Db 2880 AGCTGGGAATTCAGTGGAGCTGCAGAAAGAAACAGCAACGCTGGAAATCCAGAGATT 2939
Qy 700 GlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGly 719
Db 2940 CAATACACTTCAACTACTACAAATCTACAAATGTGGACTTTGTGTCAACACGGAAGGA 2999
Qy 720 LeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 3000 GTTTATAGCGACCTCGCCCCATTGGCACCCCGTTACCTCACCCGCAACCTG 3050
```

RESULT 7

US-10-291-583-42

; Sequence 42, Application US/10291583

; Publication No. US20030138772A1

```
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alvira, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus (
; FILE REFERENCE: US-02/735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
; LENGTH: 3122
; TYPE: DNA
; ORGANISM: 43.20
; US-10-291-583-42
```

Alignment Scores:

| | | | |
|------------------------|---------|---------------|------|
| Pred. No.: | 0 | Length: | 3122 |
| Score: | 3457.00 | Matches: | 629 |
| Percent Similarity: | 91.32% | Conservative: | 44 |
| Best Local Similarity: | 85.35% | Mismatches: | 62 |
| Query Match: | 86.66% | Indels: | 2 |
| DB: | 13 | Gaps: | 2 |

US-09-807-802A-13 (1-736) x US-10-291-583-42 (1-3122)

```
Qy 1 MetAlaAlaAspGlyTyrLeuProAspTrpLeuGluAspAsnLeuSerGluGlyIleArg 20
Db 844 ATGGCTGCCGATGGTTATCTTCAGATTGGCTCGAGGACCACTCTCTGAGGGCATTCGC 903
Qy 21 GluTrpTrpAspLeuLysProGlyAlaProLysProLysAlaAsnGlnGlnLysGlnAsp 40
Db 904 GAGTGTGGGACTTGAACCTGGAGCCCGGAAACCCAAAGCCACGACGAAACAGCAGGAC 963
Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTrpLeuGlyProPheAsnGlyLeuAsp 60
Db 964 GACGGCGGGGTCTGGTGTCTTCTGGCTACAAGTACCTCGGACCTTCAACGGACTCGAC 1023
Qy 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
Db 1024 AAGGGGAGCGCGTCAACGCGCGGACGACGCGCCCTCGAGCACGACAAAGCCTACGAC 1083
Qy 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
Db 1084 CAGCAGCTCAAGCGGGTGACATCCGTACTCTGGGTATTAATCAACGCGGACGCGAGTTT 1143
Qy 101 GlnGluArgLeuGlnGluAspThrSerPheGlyLysAsnLeuGlyArgAlaValPheGln 120
Db 1144 CAGGAGCGTCTGCAAGAAGATACGTCTTTTGGGGGCAACCTCGGGCGGAGCAGTCTTCAG 1203
Qy 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGlyAlaLysThrAlaPro 140
Db 1204 GCCAAGAAGCGGGTTCTCGAAACCTCTCGGTCTGGTTCAGGAAGCGCTTAAGACCGCTCT 1263
Qy 141 GlyLysLysArgProValGluGlnSerProGlnGlnSerProGlnGlnSerSerSerGlyIleGly 160
Db 1264 GGAAGAAGAGACTGCTGAGAGAGTCCCAACAGACGACGACTCTCTCTCGGCGATCGGC 1323
Qy 161 LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
Db 1324 AAGACAGCCAGCAGCGCGCTAAAAAGAGACTCAATTTTGTTCAGACTGCGGACTCAGAG 1383
Qy 181 SerValProAspProGlnProLeuGlyGluProAlaThrProAlaAlaValGlyPro 200
```

Db 1384 TCAGTCCCGACCCACCAACCTCTCGGAGAACCTCCAGCAGCGCCCTCAGGTCTCGGACCT 1443
Qy ThrMetAlaSerGlyGlyGlyAlaProMetAlaAspAenAenGluGlyAlaAspGly 220
Db 1444 AATACATGGCTTCAGGCGGTGGCGCTCCAATGGCAGACATAACGAAGGCGCGCAGGA 1503
Qy ValGlyAenAlaSerGlyAenTrpHisCysAspSerThrTrpLeuGlyAspArgValle 240
Db 1504 GTGGGTAATTCCTCGGAAATGGCAATTCGCAATTCACATGGCTGGGGGACAGAGTCAATC 1563
Qy ThrThrSerThrArgThrTrpAlaLeuProThrTyrAenAenHisLeuTyrLysGlnile 260
Db 1564 ACCACAGCACCCGAACTGGGCCCTGCCACCTACACCAACCACTCTACAGCAATC 1623
Qy SerSerAlaSerThrGlyAlaSer---AenAenHisTyrPheGlyTyrSerThrPro 279
Db 1624 TCCAAACGGCACCTCGGGAGGAGCACCACAGCAACACACTATTTTGGCTACAGCACCCCC 1683
Qy TrpGlyTyrPheAspAenArgPheHisCysHisPheSerProArgAspTrpGlnArg 299
Db 1684 TGGGGGTATTTTGACTTCAACAGATTCACATGTCTACTTTTACACACGCTGACTGGCAACGA 1743
Qy 300 LeuileAenAenTrpGlyPheArgProLysArgLeuAenPheLysLeuPheAenile 319
Db 1744 CTCATCAACACAAATTTGGGATTCGGGCCCAAAAGACTCACTTCAAGCTGTTCACATC 1803
Qy 320 GlnValLysGluValThrThrAenAspGlyValThrIleAlaAenAenLeuThrSer 339
Db 1804 CAGGTCMAGGAAGTCACGACGAACGAAGGACCAAGACCATCGCAATTAATCTCACACAGC 1863
Qy 340 ThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHis 359
Db 1864 ACCGTGCAGGTCTTTACGAGACTCGGAGTACCGATACCGTACGTCTAGGATCCGCTCAC 1923
Qy 360 GlnGlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeu 379
Db 1924 CAGGATGTCTGCCTCCGTTCCGGGGGAGCTTTCACGGTTCTTCAGTACCGCTATTTA 1983
Qy 380 ThrLeuAenAenGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhe 399
Db 1984 ACTTTAAACATGGAAGCAAGCCCTGGGACGTTCTCTCTTACTGTCTGGAGTATTTTC 2043
Qy 400 ProSerGlnMetLeuArgThrGlyAenAenPheThrPheSerTyrThrPheGluGluVal 419
Db 2044 CCATCCAGATGCTGAGAACCGGCAACACTTTTCACTTCACTTCACTTCACTTCACTTCACT 2103
Qy 420 PropHeHisSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAenProLeuile 439
Db 2104 CCTTTCCACAGCAGCTACCGGCACAGCCAGAGCCTGGACAGCTGATGAATCCCTCATC 2163
Qy 440 AspGlnTyrLeuTyrTyrLeuAenArgThrGlnAenGlnSerGlySerAlaGlnAenLys 459
Db 2164 GACCAGTACCTGTACTCTGTGTGAGAACGCAACG---ACTGGAACTCGAGGAGCGCAG 2220
Qy 460 AspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAenTrpLeu 479
Db 2221 ACTCTGGCATTCAGCAGCGGTCTCTAGCTCAATGGCCACACGAGCTAGAAATTTGGGTG 2280
Qy 480 ProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAenAenSer 499
Db 2281 CCGGACCTTTGTACCGGACGAGCGCTCTCCACGACCAACCAACCAACCAACCAACAGC 2340
Qy 500 AenPheThrTrpThrGlyAlaSerLysTyrAenLeuAenGlyArgGluSerIleleAen 519
Db 2341 AACTTTTGCCTGGAGCGGAGCTGCAAGTTTAAAGCTGAACGGCGGAGACTCTCTAATGAAT 2400
Qy 520 ProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGly 539
Db 2401 CCGGGCGTGCATGGCTTCCCAAGAGTACAGCAGCGCTTCTTCTTCCGAGCGGG 2460
Qy 540 ValMetIlePheGlyLysGluSerAlaGlyAlaSerAenThrAlaLeuAenAenValMet 559
Db 2461 GTCTGATTTTGGCAAGCAAGGCGGAGCAAGATGAGTGTGATTACAGCAAGTGTCTG 2520

Qy 560 IleThrAspGluGluGluLeuLysAlaThrAenProValAlaThrGluLysPheGlyThr 579
Db 2521 ATTACAGATGAGGAAGAAATCAAGGCTACCAACCCCGTGCCCAAGAAATATGGAGCA 2580
Qy 580 ValAlaValAenPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMet 599
Db 2581 GTGGCCATCAACCAACCGGCCCAATACGACGGCGCAGACCGGACTCTGTGCACCAACAG 2640
Qy 600 GlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrp 619
Db 2641 GGGGTGATTCCCGCATTCGGTGTGGCAGATAGAGACGTGTACCTGCAGGGTCCCATCTGG 2700
Qy 620 AlalysIleProHisThrAspGlyHisPheHisPheSerProLeuMetGlyGlyPheGly 639
Db 2701 GCCAAAATTCCTCACAGGACGGCACTTTTCCCGCTCTCCCTGTATGGGCGGCTTTGGA 2760
Qy 640 LeuLysAenProProProGlnIleLeuIleLysAenThrProValProAlaAenProPro 659
Db 2761 CTGNAGCACCGGCTCTCTCAATTCATCAAGAACACACCGGTTCCAGCGGAGCCCGCG 2820
Qy 660 AlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnVal 679
Db 2821 CTTACCTTCAACGAGGCAAGCTGAACCTCTTTTATCATCAGCAGTACAGCACCGGACGTC 2880
Qy 680 SerValGluIleGluTTPGluLeuGlnLysGluAenSerLysArgTTPAenProGluVal 699
Db 2881 AGCGTGAATTCGAGTGGGAGCTGCAGAAAGAAACAGCAACCGTGGATCCAGAGATT 2940
Qy 700 GlnTyrThrSerAenTyrAlaLysSerAlaAenValAspPheThrValAspAenAenGly 719
Db 2941 CAATACACTTCCAATCTACAAATCTACAAATCTGGACTTTTGTCTCAACACCGMAGGA 3000
Qy 720 LeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 3001 GTTTATAGCAGCGCTCGCCCATTTGGCAGCCGTTTACCTCACCCGCAACCTG 3051

RESULT 8

US-10-291-583-27
; Sequence 27, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alvira, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; FILE REFERENCE: UPN-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 27
; LENGTH: 3128
; TYPE: DNA
; ORGANISM: 42.8
US-10-291-583-27

Alignment Scores:

| | | | |
|------------------------|---------|---------------|------|
| Pred. No.: | 0 | Length: | 3128 |
| Score: | 3453.00 | Matches: | 628 |
| Percent Similarity: | 91.13% | Conservative: | 45 |
| Best Local Similarity: | 89.09% | Mismatches: | 63 |
| Query Match: | 86.56% | Indels: | 2 |
| DB: | 13 | Gaps: | 2 |

| | | | | | |
|---|---|----|------|---|------|
| US-09-807-802A-13 (1-736) x US-10-291-583-27 (1-3128) | | Db | 1864 | AGCAGATTTCAGGTCTTTACGGACTCGGAATACACAGCTCCCGTACGTCCTCGGCTCTCGC | 1923 |
| Qy | 1 MetAlaAlaAspGlyTyrLeuProAspTrpLeuGluAspAenLeuSerGluGlyLeuArg | Qy | 359 | HisGlnGlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGlyTyr | 378 |
| Db | 844 ATGGCTCGCGATGGTATCTTCCAGATTGGCTCGAGGACAACCTCTCTGAGGGCATTCGC | Db | 1924 | CACAGGGCTGCTGCTCGTTCCTGGCGGAGAGTCTTCATGATTCCTCAGTACGGGTAC | 1983 |
| Qy | 21 GluTrpTrpAspLeuLeuProGlyAlaProLysProLysAlaAenGlnGlnLysGlnAsp | Qy | 379 | LeuThrLeuAenAenGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyr | 398 |
| Db | 904 GAGTGGTGGACTTGAACCTGGAGCCCGGAAACCCAAAGCCAAACAGCAAGCAGGAC | Db | 1984 | CTGACTCTGAACAACGGCAGTCAGGCGGTGGGCGTCTCTCTCTCTCTCTCTCTCTCTCT | 2043 |
| Qy | 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAenGlyLeuAsp | Qy | 399 | PheProSerGlnMetLeuArgThrGlyAenAspPheThrPheSerTyrThrPheGluGlu | 418 |
| Db | 964 GACGGCCGGGTCTGGTCTCTCTGGCTACAGTACCTCGGACCTCTCAACGGACTCGAC | Db | 2044 | TTTCTCTCTCAATGCTGAGAACGGGCAACAACTTTGAGTTTCAGTACCTACCTTTGAGAC | 2103 |
| Qy | 61 LysGlyGluProValAenAlaAspAlaAlaAlaLeuGluHisAspLysAlaTyrAsp | Qy | 419 | ValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeu | 438 |
| Db | 1024 AAGGGGAGCCCGTCAACCGCGGAGCAGCAGCGCCCTCGAGCAGCAGCAAGCCCTACGAC | Db | 2104 | GTGCTTTTTCACAGCAGCTACGGCAGCAGCAAGCCCTGACCGGCTGATGAAACCCCTC | 2163 |
| Qy | 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAenHisAlaAspAlaGluPhe | Qy | 439 | IleAspGlnTyrLeuTyrTyrLeuAenArgThrGlnAenGlnSerGlySerAlaGlnAen | 458 |
| Db | 1084 CAGCAGCTCAAGCGGGTGACAACTCCGTACCTGGGTATACCAACCGCGCCGAGTTT | Db | 2164 | ATCGACCAGTACCTGTACTCTCGGACTCTCGGACTCAGTCCACGGGAGGTACCGCAGGA | 2223 |
| Qy | 101 GlnGluArgLeuGlnGluAspThrSerPheGlyAenLeuGlyArgAlaValPheGln | Qy | 459 | LysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAenTrp | 478 |
| Db | 1144 CAGGAGCGCTCTCAAGAGATACGTCTTTGGGGGCAACCTCGGCGCAGCAGTCTTCAG | Db | 2224 | CAGCAGTTGCTATTTTCTCAGCGCGGCGCTTAATACATGTCTCGGCTCAGGCAAAATCGG | 2283 |
| Qy | 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGlyAlaLysThrAlaPro | Qy | 479 | LeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAenAen | 498 |
| Db | 1204 GCCAAGAGCGGGTTCTCGAACCTCTCGGTCTGGTTGAGGAAGCGCTTAAGCCGGCTCT | Db | 2284 | CTACCGGGGCGCTGTACCGCGCAGCAACGGCTCTCCACGACACGTCTCGCAAAATAACA | 2343 |
| Qy | 141 GlyLysLysArgProValGluGlnSerProGln---GluProAspSerSerSerGlyLe | Qy | 499 | SerAsnPheThrTrpThrGlyAlaSerLysTyrAenLeuAenGlyArgGluSerIleIle | 518 |
| Db | 1264 GGAAAGAGAGACCGGTAGAGCCATCACCCCGCGGTTCTCCAGACTCTCTACGGGCATC | Db | 2344 | AGCAACTTTGCTTGGACCGGTGCACCAAGTATCATCTGAATGGCAGAGACTCTCTCGTA | 2403 |
| Qy | 160 GlyLysThrGlyGlnGlnProAlaLysLysArgLeuAenPheGlyGlnThrGlyAspSer | Qy | 519 | AsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSer | 538 |
| Db | 1324 GGCAGACAGCCAGCAGCCGCGGAAAGAGACTCAACTTTGGCGCAGACTGGCGGACTCA | Db | 2404 | ATTCGGGTCTGCTATGCGCAACGCAAGCAGCAGCAAGAGCGATTTTTTTCATCCAGC | 2463 |
| Qy | 180 GluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGly | Qy | 539 | GlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnVal | 558 |
| Db | 1384 GAGTCAGTGGCCGACCTTCAACCAATCGAGAACCCCGCGCAGCGCTCTGTCTGGGA | Db | 2464 | GGAGTCTTGATGTTTGGGAAACAGGGAGCTGGAAAGACCAACCGTGGACTATAGCAGGTT | 2523 |
| Qy | 200 ProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAenAenGluGlyAlaAsp | Qy | 559 | MetIleThrAspGluGluGluLysAlaThrAsnProValAlaThrGluLysArgPheGly | 578 |
| Db | 1444 TCTGTACAATGGCTGCGAGCGGTGGCGCTCCAATGGCAGACAATAACGAAGCGCGAC | Db | 2524 | ATGCTAACCAAGTGGAGAAATCAAAACCAACCAACCAAGTGGCCACAGCAAGATACGGC | 2583 |
| Qy | 220 GlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgVal | Qy | 579 | ThrValAlaValAenPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAla | 598 |
| Db | 1504 GGAGTGGGTAGTTCTCTCAGGAAATTTGGCATTTGCGATTCCACATGGCTGGCGCAGAGTC | Db | 2584 | GTGGTGGCCGATACCTGCAACACAGCAAAACCGCCCTCTATTGTAGGGGGCGTCAACAGT | 2643 |
| Qy | 240 IleThrThrSerThrArgThrTrpAlaLeuProThrTyrAenAenHisLeuTyrLysGln | Qy | 599 | MetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIle | 618 |
| Db | 1564 ATCACACAGCAGCCCGAACTGGGCCCTTCCACCTACCAACCAACCACTCTACAGCA | Db | 2644 | CAAGGAGCTTACTTGGCATGGTCTGGCAGAACCGGAGCGTGTACTCTGAGGGTCTCTATC | 2703 |
| Qy | 260 IleSerSerAlaSerThrGlyAlaSer---AenAspAenHisTyrPheGlyTyrSerThr | Qy | 619 | TrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPhe | 638 |
| Db | 1624 ATCTTCAACGGGACATCGGGAGGAGCACCACCAACGACACACTACTTCTGGCTACAGC | Db | 2704 | TGGGCCCAAGATTCTCACAACGACGCACTTTTCATCTCTCGCGCTGATGGGAGGCTTT | 2763 |
| Qy | 279 ProTrpGlyTyrPheAspAenArgPheHisCysHisPheSerProArgAspTrpGln | Qy | 639 | GlyLeuLysAenProProGlnIleLeuLysAenThrProValProAlaAenPro | 658 |
| Db | 1684 CCCTGGGGGTATTTGATTTTAAACAGATTCCACTTGCCTTCTCACCACGTGACTGGCAG | Db | 2764 | GGACTGAAACACCGCCCTCTCTCAGATCTGATTAAGAAATACCTGTCTCCCGCGATCCT | 2823 |
| Qy | 299 ArgLeuIleAenAenAsnTrpGlyPheArgProLysArgLeuAenPheLysLeuPheAen | Qy | 659 | ProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGln | 678 |
| Db | 1744 CGACTCATCAACCACTCTGGGATTCGGGCCCAAGAGACTCAACTTCAAGCTCTTCAAC | Db | 2824 | CCAACTACCTTTCAGTCAAGCAAGCTGGCGTCTGTTTCATCAGCAGTACAGCAGCGGACAG | 2883 |
| Qy | 319 IleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAenAenLeuThr | Qy | 679 | ValSerValGluIleGluTrpGluLeuGlnLysGlnAenSerLysArgTrpAsnProGlu | 698 |
| Db | 1804 ATCCAGGTCAAGGAGGTCAAGCAGAAATGAAGGACCAAGGACCAATCGCCAAATACCTTACC | Db | 2884 | GTACAGCTGGAATTTGAATGGAGCTGTCAGAAAGAGAACAGCAGCGCTGGAACCCAGAG | 2943 |
| Qy | 339 SerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAla | Qy | 699 | ValGlnTyrThrSerAenTyrAlaLysSerAlaAenValAspPheThrValAspAenAen | 718 |
| | | Db | 2944 | ATTGATGATATCTTCCAACTACTACAAATCTACAAATGTGAGCTTTGCTGTCACTACTGAG | 3003 |

Qy 719 GlyLeuTyrThrGluProArgProTleGlyThrArgTyrLeuThrArgProLeu 736
Db 3004 GGTACTTATTCAGAGCTCGCCCATTTGGACCCGTTACCTCACCGTAACCTG 3057

RESULT 9
US-10-291-583-26
; Sequence 26, Application US/10291583
; Publication NO. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alvira, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; FILE REFERENCE: UPN-02735USA
; CURRENT FILING DATE: 2002-11-12
; PRIOR FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26
; LENGTH: 3075
; TYPE: DNA
; ORGANISM: new AAV serotype, clone H2
US-10-291-583-26

Alignment Scores:
Pred. No.: 0 Length: 3075
Score: 3451.50 Matches: 622
Percent Similarity: 91.17% Conservative: 49
Best Local Similarity: 84.51% Mismatches: 64
Query Match: 86.53% Indels: 1
DB: 13 Gaps: 1

US-09-807-802A-13 (1-736) x US-10-291-583-26 (1-3075)

Qy 1 MetAlaAlaAspGlyTyrLeuProAspTrpLeuGluAspAsnLeuSerGluGlyIleArg 20
Db 803 ATGGCTGCCGATGGTATCTCCAGATTGGCTCGAGGACACTCTCTCTGAGGGATAAGA 862

Qy 21 GluTrpTrpAspLeuLysProGlyAlaProLysProLysAlaAsnGlnGlnLysGlnAsp 40
Db 863 CAGTGTGGAAGCTCAAACTGGCCACCACCACCACCACCACCACCACCACCACCACCACC 922

Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
Db 923 GACAGCAGGGCTCTTGCTTCTCGGTACAAAGTACCTCGGACCTTCAACGGACTCGAC 982

Qy 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
Db 983 AAGGGGGAGCCGTCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1042

Qy 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
Db 1043 CGGCAGCTCGACAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1102

Qy 101 GlnGluArgLeuGlnGluAspTrpSerPheGlyAsnLeuGlyArgAlaValPheGln 120
Db 1103 CAGGAGCGCTTAAAGAAGATAGTCTTTTGGGGGCAACTCTGGAGGAGGAGGAGGAGG 1162

Qy 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGlyAlaLysThrAlaPro 140
Db 1163 GCCAAAAGAGGGTCTTGAACCTCTGGGCTGGTGGAGAACCTGTTAAGACGGCTCGG 1222

Qy 141 GlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerGlyIleGly 160

Db 1223 GGAAGAGAGCGGCTAGAGCACTCTCTGTGGAGCGAGCACTCTCTCTGGGAAACCGGA 1282
Qy 161 LysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGlu 180
Db 1283 AAGCGGGCCAGCGGCTGCAAGAAAAGATTAATTTTGGTCAGACTGGAGAGCGCAGAC 1342

Qy 181 SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro 200
Db 1343 TCCGTACTCTACCCCGAGCTCTCGGACAGCCAGCAGCGCCCTCTGTGCTGGGATCT 1402

Qy 201 ThrThrMetAlaSerGlyGlyGlyValaProMetAlaAspAsnAsnGluGlyValaAspGly 220
Db 1403 ACTACATGCTACAGCAGCTGGCGCACCAATGCGACACAATAACAGGGGTGCGGATGA 1462

Qy 221 ValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIle 240
Db 1463 GTGGGTAAATCTCTCAGGAAATTTGGCATTTGGCATTTCCCAATGGCTGGGCGCAGAGTCA 1522

Qy 241 ThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIle 260
Db 1523 ACCACGACGACCCGAGCTGGGCGCTCCCAACATACACCAACCACTCTACAGCAATC 1582

Qy 261 SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrp 280
Db 1583 TCCAGCCCAATCA---GGAGCCAGCAACGACCAACCACTACTTTGGCTACAGCACCCTCTG 1639

Qy 281 GlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu 300
Db 1640 GGGTATTTTGACTTCAACAGATTCACCTGCCACTTTTCCACCACCTGAGTGGCAAGACTC 1699

Qy 301 IleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGln 320
Db 1700 ATCAACCAACAATCGGGGATTCGGGCCCAAAAGACTCAACTTCAAGCTCTTTAATATTCA 1759

Qy 321 VallysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThr 340
Db 1760 GTCAAGAGGTCACGCGAGATGACGGTACGAGCAGGATTTGCCAATACTTACCAGCAGC 1819

Qy 341 ValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGln 360
Db 1820 GTTCAGGTGTTTACTGACTCGGAGTACGAGTCCCGTCCGTCCTGGGCTCGGCGCATCAA 1879

Qy 361 GlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThr 380
Db 1880 GGAATGCTTCCCGCGTTCACGAGCGAGCTCTTCATGGTCCACAGATGAGTACCTCACC 1939

Qy 381 LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 400
Db 1940 CTGAACAACGGGAGTACGCGGTAGGAGCTCTTCTTTTACTGCTGCTGGAGTACTTCTCT 1999

Qy 401 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 420
Db 2000 TCTCAGATGCTGCGTACTGSAACAACTTTTCAGTTTCAGTACACTTTTGAAGAGCTGCT 2059

Qy 421 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 440
Db 2060 TTCCACAGCAGTACGCTCACAGCAGAGTCTGGATCGCTGATGATGATCTCTGTATGAC 2119

Qy 441 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAsp 460
Db 2120 CAGTACCTGTATTATCTGACACACAAACAAATAGTGGAACTCTTTCAGCAGCTCTCGG 2179

Qy 461 LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysProLysAsnTrpLeuPro 480
Db 2180 CTACTGTTTAGCAGCTGGACCAACCACTGCTCTTCAAGCTTAAAGACTGGCTGCTCT 2239

Qy 481 GlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsn 500
Db 2240 GGACCTTCTACAGCAGCAGCGCTGTCTGAAAACAGGCAACGACCAACCAACAGCAAC 2299

Qy 501 PheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnPro 520

Db 2300 TTTCCTCGAGCTGACGATCAAGATATCATCTAAATGCGCGGACTCGTTGGTTAAATCCA 2359
Qy GlyThrAlaMetAlaSerHisLeuAspAspGluAspLysPhePheProMetSerGlyVal 540
Db GGACCAAGTATGGCCAGTCACAGATGACGAGAAAGATTTTCCCATGTCATGGAAACC 2419
Qy MetLeuPheGlyLeuGluSerAlaGlyAlaSerHisThrAlaLeuAspAsnValMetIle 560
Db CTGATATTTGTAACCAAGCAACAAATGCCAACGACGCGGATTTGGAAATGTCTATGATT 2479
Qy ThrAspGluGluGluLeuLeuAlaThrAsnProValAlaThrGluArgPheGlyThrVal 580
Db ACAGATGAAGAGAAATCAGGCCACCAATCCCGTGGCTACGGAGCAGATCGGAGACTGTG 2539
Qy AlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGly 600
Db TCAATAATTTGCAAACTCAACACTCGTCCACTACTGGAATGTCTCAATCGCCAGGA 2599
Qy AlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAla 620
Db GCGTTACCTGTGTGGTGGCAGATCGAGACGTGTACTTGCAGGAGACCATTTGGGCC 2659
Qy LysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeu 640
Db AAGATTCTCTACACCGATGGAACATTTTCATCTCTCCACTGATGGGAGGTTTGGACTC 2719
Qy LysAsnProProGlnIleLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 660
Db AAACACCGCTCTCTCAGATCATGATCAAAACACTCCCGTTCAGCCCATCTCCCNCA 2779
Qy GluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSer 680
Db AACTTCAGTTCTGCCAAGTTTCTTCTTCTATCACACAGTATTCACCGGACAGGTCAAG 2839
Qy ValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTyrAsnProGluValGln 700
Db GTGGAGATCGAGTGGAGCTGCGAAGGAGAGAACAGCAACGCTGGAATCCCGAAATTCAG 2899
Qy TyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeu 720
Db TACACTTCCAACTACAACAGTCTGTAAATGTGGACTTTACTGTGGACACTAATGGTGTG 2959
Qy TyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db TATTCAGAGCTCGCCCATTTGGCACCAGATACCTGACTCGTAAATCTG 3007

RESULT 10
US-10-291-583-59
; Sequence 59, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alvira, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; FILE REFERENCE: UPN-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59
; LENGTH: 3129
; TYPE: DNA
; ORGANISM: 44.2

US-10-291-583-59
Alignment Scores:
Pred. No.: 0 Length: 3129
Score: 3448.00 Matches: 627
Percent Similarity: 91.19% Conservative: 46
Best Local Similarity: 84.96% Mismatches: 63
Query Match: 86.44% Indels: 2
DB: 13 Gaps: 2

US-09-807-802A-13 (1-736) x US-10-291-583-59 (1-3129)

Qy 1 MetAlaAlaAspGlyTyrLeuProAspTrpLeuGluAspAsnLeuSerGluGlyIleArg 20
Db 845 ATGGCTGCCGATGTTATCTTCAGATTGGCTCGAGACAACCTCTCTGAGGGCATTCGC 904
Qy 21 GluTrpTrpAspLeuLysProGlyAlaProLysProLysAlaAsnGlnGlnLysGlnAsp 40
Db 905 GAGTGTGGGACTTGAACCTTGAGGCCCGGAAACCCAAAGCCAAAGAGAGGAC 964
Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
Db 965 GACGCCCGGGGTCTGGTCTTCTGGCTACAGTACTCTCGGACCTTCAACGGACTCGAC 1024
Qy 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
Db 1025 AAGGGGAGCCCTCAACGGCGGACGACGAGCGCCCTCGAGCACGACAAGGCTTACGAC 1084
Qy 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
Db 1085 CAGCAGCTCAAGCGGGTGACAAATCCGTACTCTCGGTATTAACCAACGCGCGAGGTTT 1144
Qy 101 GluGluArgLeuGlnGluAspThrSerPheGlyValAsnLeuGlyArgAlaValPheGln 120
Db 1145 CAGGAGGCTCTGCAAGAGATACGTCTTTTGGGGCAACCTCGGGCGAGCAGTCTTCAG 1204
Qy 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGlyAlaLysThrAlaPro 140
Db 1205 GCCAAGAAGCGGGTCTCGAACCTCTCGTCTGGTGGAGAAAGCGCTTAAGACGGCTCT 1264
Qy 141 GlyLysLysArgProValGluGlnSerProGln---GluProAspSerSerGlyIle 159
Db 1265 GGAAGAAGAGACCGGTAGAGCCATCAACCCAGCGTTCTCCAGACTCTCTACGGGATC 1324
Qy 160 GlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSer 179
Db 1325 GGCAGAAAGCCGACGACGCGCGGAAAGAGACTCACTTTGGGACAGCTGGCGACTCA 1384
Qy 180 GluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaValGly 199
Db 1385 GAGTCAGTGGCCGACCTCAACCAATCGAGAACCCCGCGAGCCCTCTCTGGTCTGGGA 1444
Qy 200 ProThrThrMetAlaSerGlyGlyValProMetAlaAspAsnGluGlyAlaAsp 219
Db 1445 TCTGGTACAATGGCTGACGCGGTGGCGCTCCAAATGGCAGACATAACGAAAGGCCGAC 1504
Qy 220 GlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgVal 239
Db 1505 GGAGTGGGTAGTTCTCTCAGGAAATGGCATTTGGCATTTCCATGGCTGGCGACAGATC 1564
Qy 240 IleThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGln 259
Db 1565 ATCACCACGACGACCCGAACTCGGGCCCTCCCACTTACCAACCAACCACTCTACAGCNA 1624
Qy 260 IleSerSerAlaSerThrGlyAlaSer---AsnAspAsnHisTyrPheGlyTyrSerThr 278
Db 1625 ATCTCCAACGGGACTTCGGAGGAGGAGCACCACGACCAACCACTTCTTGGCTACGAC 1684
Qy 279 ProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGln 298
Db 1685 CCCTGGGGGTATTTGACTTTAAGATTCACAGATTCCTGACCTTCTCACCACGCTGGCAG 1744
Qy 299 ArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsn 318

Db 1745 CGACTCATCAACAACTGGGGATTCCGGCCCAAGAGACTCAACTTCAAGCTCTTCAAC 1804
Qy 319 IleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThr 338
Db 1805 ATCCAGGTCAAGAGGTCAAGCAGATGAAGGACCAAGACCATCGCCAATAACCTTACC 1864
Qy 339 SerThrValGlnValPheSerAspSerGluThrGlnLeuProTyrValLeuGlySerAla 358
Db 1865 AGCAGGATTTCAGGCTCTTTACGAGCTCGGAATACCAAGCTCCCGTACGCTCTCGGCTCGG 1924
Qy 359 HisGlnGlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGlyTyr 378
Db 1925 CACCAGGGTGGCTGGCTCGGTCCTCCGGCGGAGCTTTCATGATTCCTCAGTACGGGTAC 1984
Qy 379 LeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyr 398
Db 1985 CTGACTCTGAACAAATGGCAGTCAAGCGGTGGGCGGTTCCTCTTACTGCTCGGAGTAC 2044
Qy 399 PheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGlu 418
Db 2045 TTTCCTTCTCAATGCTGAGAACGGGCAACAACTTTGAGTTTCAGCTTACCAAGTTTGAAGAC 2104
Qy 419 ValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeu 438
Db 2105 GTGCCCTTTTCACAGCAGCTACGGCCACAGCCAAAGCTGGACCGGCTGATGAAACCCCTC 2164
Qy 439 IleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsn 458
Db 2165 ATGACCACTGACTGACTGCTGCTGAGCTCAGTCCACGGGAGGTACCCGAGAACT 2224
Qy 459 LysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrp 478
Db 2225 CAGCAGTTGCTATTTCTCAGCGCGGGCCCTTAATCAATGTCGGCTCAGGCCCAAACTGG 2284
Qy 479 LeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 498
Db 2285 CTACCCGGGCGCTGCTACCGGAGCAACCGGCTCTCCACACACTGTCCGCAAAATAACAC 2344
Qy 499 SerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIle 518
Db 2345 AGCAACTTTGCTGGACCGGTGCGCACCAAGTATCATCTGAATGGCAGAGACTCTCTGTA 2404
Qy 519 AsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPheProMetSer 538
Db 2405 AATCCCGGTGCTGCTATGGCAATCCCAAGGACGAGAGAGGATTTTTCGTCCAGC 2464
Qy 539 GlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnVal 558
Db 2465 GGAGTCTTATGTTTGGAAACAGGGAGCTGGAAAGACCAACGCTGAGCTATAGCAGCTT 2524
Qy 559 MetIleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGly 578
Db 2525 ATGCTAACCACTGAGGAAGAAATTAACCCACCAACCAAGTGGCCACAGAACAGTACGGC 2584
Qy 579 ThrValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAla 598
Db 2585 GTGTGGCCGATACCTGCAACAGCAAAACCGCGCTCTTATGTTAGGGCCGCTCAACAGT 2644
Qy 599 MetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIle 618
Db 2645 CAAGGACCTTACTTGCATGGTCTGGCAGAACCGGACGCTGTACCTGCGAGGCTCTATC 2704
Qy 619 TrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPhe 638
Db 2705 TGGGCCAAGATTCTCACAGGACGGAAACTTTTCATCCCTCGCGCTGTATGGGAGGCTTT 2764
Qy 639 GlyLeuLysAsnProProGlnIleLysLeuIleLysAsnThrProValProAlaAsnPro 658
Db 2765 GGACTGAACACACCCGCTCTCTCAGATCTCTGATTAAGANTACCTGTTCCCGCGGATCCT 2824
Qy 659 ProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGln 678

Db 2825 CCAACTACCTTCACTCAAGCTAAGCTGGCTGCTTCATCAGCAGTACAGCACCGGACAG 2884
Qy 679 ValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGlu 698
Db 2885 GTCAGCGTGGAAATTAATGGAGCTGCAGAAAGAAACAGCAAAACGCTGGAAACCCAGAG 2944
Qy 699 ValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsn 718
Db 2945 ATTCAATACACTTCCAACTACTACAATCTACAATGTGGACTTTGCTGTTAACACAGAT 3004
Qy 719 GlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 3005 GGCACCTTATTCTGAGCTCGCCCATCGCACCGCTTACCTCAGCCGCTAATCTG 3058

RESULT 11

US-10-291-583-25/c

; Sequence 25, Application US/10291583

; Publication No. US20030138772A1

; GENERAL INFORMATION:

; APPLICANT: Gao, Guangping

; APPLICANT: Wilson, James M.

; APPLICANT: Alvira, Mauricio

; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus

; FILE REFERENCE: UPN-02735USA

; CURRENT APPLICATION NUMBER: US/10/291,583

; CURRENT FILING DATE: 2002-11-12

; PRIOR APPLICATION NUMBER: US 60/350,607

; PRIOR FILING DATE: 2001-11-13

; PRIOR APPLICATION NUMBER: US 60/341,117

; PRIOR FILING DATE: 2001-12-17

; PRIOR APPLICATION NUMBER: US 60/377,066

; PRIOR FILING DATE: 2002-05-01

; PRIOR APPLICATION NUMBER: US 60/386,675

; PRIOR FILING DATE: 2002-06-05

; NUMBER OF SEQ ID NOS: 120

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 25

; TYPE: DNA

; ORGANISM: new AAV serotype, clone H6

US-10-291-583-25

Alignment Scores:

| | | | |
|------------------------|---------|---------------|------|
| Pred. No.: | 0 | Length: | 3142 |
| Score: | 3446.50 | Matches: | 622 |
| Percent Similarity: | 91.17% | Conservative: | 49 |
| Best Local Similarity: | 84.51% | Mismatches: | 64 |
| Query Match: | 86.40% | Indels: | 1 |
| DB: | 13 | Gaps: | 1 |

US-09-807-802A-13 (1-736) x US-10-291-583-25 (1-3142)

Qy 1 MetalAlaAspGlyTyrLeuProAspTrpLeuGluAspAsnLeuSerGluGlyIleArg 20
Db 2339 ATGGCTGCCGATGGTATCTTCCAGATGGCTCGAGGACACTCTCTCTGAGGAATAAGA 2280
Qy 21 GluTrpTrpAspLeuLysProGlyAlaProLysProLysAlaAsnGlnLysGlnAsp 40
Db 2279 CAGTGGTGGAGCTCAACCTGGCCCAACCCAGAGCCGCGCATATAGGAC 2220
Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
Db 2219 GACAGCAGGGTCTTGTGCTTCTGGGTACAGTACCTCGGACCTTCAACGAGCTCGAC 2160
Qy 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
Db 2159 AAGGGGAGCCGCTCAACGAGGACAGCGCGCGGCTCGAGCAGCAAGGCTTACGAC 2100
Qy 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
Db 2099 CGGCAGCTCGACAGCGGAGCAACCCGTACCTCAAGTACACCAACCGCCGCGAGGTTT 2040

| | | | |
|----|------|--|------|
| Qy | 101 | GlnGluArgLeuGlnGlnGluAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln | 120 |
| Db | 2039 | CAGAGCGCCCTTAAAGAAAGATACGCTCTTTTGGGGGCGCAACCTCGGACAGAGAGTCTTCCAG | 1980 |
| Qy | 121 | AlaIysIysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaIaValThrAlaPro | 140 |
| Db | 1979 | CGCAAAAGAGGGTCTTTGAACCTCTTGCGCTGTGTGAGGAGCCTGTTAAGACGGCTCCG | 1920 |
| Qy | 141 | GlyIysIysArgProValGluGlnSerProGlnGluProAspSerSerSerGlyIleGly | 160 |
| Db | 1919 | GGAAAAAGAGCGCGTAGAGCACTCTCTCTGTGGAGCCAGACTCTCTCTCGGGAACCGGA | 1860 |
| Qy | 161 | LysThrGlyGlnGlnProAlaIaIysIysArgLeuAsnPheGlyGlnThrGlyAspSerGlu | 180 |
| Db | 1859 | AAAGCGGGCCAGCAGCCCTGCAAGAAAGATTAATTTTGGTCAGACTGGAGACAGAC | 1800 |
| Qy | 181 | SerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyPro | 200 |
| Db | 1799 | TCCGTACTGACCCCGCCAGCCCTCTCGGACAGCCACAGAGCGCCCTCTGGTCTGGGATCT | 1740 |
| Qy | 201 | ThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGly | 220 |
| Db | 1739 | ACTACAATGGCTACAGGCAGTGGCGCCACCAATGGCAGACAATAACAGGGGTGCCGATGGA | 1680 |
| Qy | 221 | ValGlyAsnAlaSerGlyAsnThrHisCysAspSerThrTrpLeuGlyAspArgValIle | 240 |
| Db | 1679 | GTGGGTATTCCTCAGGAAATGGCAATTGGCAATTGCCAATTCGCTGGGCGCAGAGTCAATC | 1620 |
| Qy | 241 | ThrThrSerThrArgThrTrpAlaLeuProThrTyrrhenAsnHisLeuTyrrLysGlnIle | 260 |
| Db | 1619 | ACCACAGCACCCGAACCTGGGCGCTGCCACATACAAACCACTCTTCACAGCAAAATC | 1560 |
| Qy | 261 | SerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrrPheGlyTyrrSerThrProTyr | 280 |
| Db | 1559 | TCCAGGCCAATCA--GGAGCGAGCAGACGACAACTACTTTGGCTACAGCACCCCTCTGG | 1503 |
| Qy | 281 | GlyTyrrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeu | 300 |
| Db | 1502 | GGGTATTTTGACTTCAACAGATTCCACTGCCACATTTTCCACCGTGACTGCAAAAGACTC | 1443 |
| Qy | 301 | IleAsnAsnAsnTrpGlyPheArgProIysArgLeuAsnPheIysLeuPheAsnIleGln | 320 |
| Db | 1442 | GTCAACAAACACCGGGGATTCGCGCCCAAAAGACTCAACTTCAAGCTCTTTATATTATTCAA | 1383 |
| Qy | 321 | ValIysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThr | 340 |
| Db | 1382 | GTCAAGAGGTTCAGCAGATGACGGTACGACGACGATGGCCATATACCTTACAGGACGC | 1323 |
| Qy | 341 | ValGlnValPheSerAspSerGluTyrrGlnLeuProTyrrValLeuGlySerAlaHisGln | 360 |
| Db | 1322 | GTTCAGGTGTTTACTGACTCCGAGTACCAGCTCCCGTACGTCTCTGGGCTCGGCGCATCAA | 1263 |
| Qy | 361 | GlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrrGlyTyrrLeuThr | 380 |
| Db | 1262 | GGATGCTCTCCCGCGTTCACAGCGGACGCTCTTCATGGTCCACAGTATGATACCTCACC | 1203 |
| Qy | 381 | LeuAsnAsnGlySerGlnAlaValIysArgSerSerPheTyrrCysLeuGluTyrrPhePro | 400 |
| Db | 1202 | CTGAACAAACGGGAGTACGCGGTGAGGACGCTCTCTCTTTTACTGCCCTGGAGTACTTTCT | 1143 |
| Qy | 401 | SerGlnMetLeuArgThrGlyIysAsnAsnPheThrPheSerTyrrThrPheGluGluValPro | 420 |
| Db | 1142 | TCTCAGATGCTGCTACTTGGAAACAACCTTTCAGTTTCAGTACACTTTTGAAGACGTCGCT | 1083 |
| Qy | 421 | PheHisSerSerTyrrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp | 440 |
| Db | 1082 | TTCCACAGCAGCTACGCTCAGCCAGAGTCTGGATCGGCTGATGATGAATCTCTCTGATCGAC | 1023 |
| Qy | 441 | GlnTyrrLeuTyrrTyrrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAsp | 460 |
| Db | 1022 | CAGTACCTGTATATCTGAACAAGACACAACAAATAGTGGAACTCTTCAGCAGTCTCGG | 963 |
| Qy | 461 | LeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuPro | 480 |

[illegible]

RESULT 12

```

US-10-291-583-41
; Sequence 41, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alvira, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; FILE OF INVENTION: Sequences and Isolating No. US20030138772A1el Sequences Identif.
; FILE REFERENCE: UFN-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066

```

;; PRIOR FILING DATE: 2002-05-01
;; PRIOR APPLICATION NUMBER: US 60/386,675
;; PRIOR FILING DATE: 2002-06-05
;; NUMBER OF SEQ ID NOS: 120
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 41
;; LENGTH: 3123
;; TYPE: DNA
;; ORGANISM: 43.12
US-10-291-583-41

Alignment Scores:

Pred. No.: 0 Length: 3123
Score: 3445.00 Matches: 625
Percent Similarity: 91.19% Conservative: 48
Best Local Similarity: 84.69% Mismatches: 63
Query Match: 86.36% Indels: 2
DB: 13 Gaps: 2

US-09-807-802a-13 (1-736) x US-10-291-583-41 (1-3123)

QY 1 MetAlaAlaAspGlyTyrLeuProAspTyrLeuGluAspAsnLeuSerGluGlyIleArg 20
DB 839 ATGGCTGCCGATGGTATCTCCAGATTGGCTTGAGGCAACCTCTCTGAGGSCATTGCG 898
QY 21 GluTyrTyrAspLeuLysProGlyAlaProLysProLysAlaAsnGlnLysGlnAsp 40
DB 899 GAGTGTGGGACCTGAACCTCGAGCCCGCAAAACCAAGCCCAACAGCAAGAGCAGGAC 958
QY 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
DB 959 GACGGCCGGGGTCTGGTCTCTTGGCTACAGTACCTCGGACCTTTCAACGGACTCGAC 1018
QY 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
DB 1019 AAGGGGAGCCGCTCAACGGCGGAGCGAGCGGCCCTCGAGCAGCAAGGCTTACGAC 1078
QY 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
DB 1079 CAGCAGCTCAAGCGGGTGACATCCGTACCTCGGTGATTAACACGCGCAGCCGAGTTT 1138
QY 101 GlnGluArgLeuGlnGluAspThrPheGlyGlyAsnLeuGlyArgAlaValPheGln 120
DB 1139 CAGGAGGCTGCAAGAGNAGTCTCTTTTGGGGGCAACTCTGGGCGAGCAGTCTTCCAG 1198
QY 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaPro 140
DB 1199 GCCAAGAAGCGGTTCTCGAACCTCTCGGTCTGGTTGAGGAAGCGCTAAGACGGCTCCT 1258
QY 141 GlyLysLysArgProValGlnGlnSerProGln--GluProAspSerSerSerGlyIle 159
DB 1259 GGAAGAAGAGACCGGTAGAGCCATCACCTCAGCGTTCCCGGACTCTCCACGGGCATC 1318
QY 160 GlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSer 179
DB 1319 GGCAGAAAGCCACAGCCCGGAGAAAGAGACTGAACTTTGGGCGAGACTGGCGACTCG 1378
QY 180 GluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaValGly 199
DB 1379 GAGTCAGTCCCGGACCTCAACCAATCGGAGAACCCAGCAGCGCCCTCTGGTCTGGGA 1438
QY 200 ProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnGlnGlyAlaAsp 219
DB 1439 TCTGGTACATGGCTGAGCGGTGGCGCTCCAAATGGCAGACAAATAACGAAGCGCCGAC 1498
QY 220 GlyValGlyAsnAlaSerGlyAsnTyrHisCysAspSerThrTyrLeuGlyAspArgVal 239
DB 1499 GGAGTGGTAGTTCCTCAGGAATTTGGCATTTGGATTCCACATGGCTGGGCGCAGAGTC 1558
QY 240 IleThrThrSerThrArgThrTyrAlaLeuProThrTyrAsnAsnHisLeuTyrLysGln 259
DB 1559 ATCACCACGACGACCCGACCTGGGCGCTGCCACCTTACCAACCAACCTCTCTACAGCAA 1618

QY 260 IleSerSerAlaSerThrGlyAlaSer--AsnAspAsnHisTyrPheGlyTyrSerThr 278
DB 1619 ATCTCAACCGGACATCGGAGGAAGCACTAAACGACCAACCTACTTTGGCTACAGCAC 1678
QY 279 ProTyrGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTyrGln 298
DB 1679 CCCTGGGGGTATTTTGACTTCAACAGATTCCACTGCCACTTCTCACCAAGCTGGCAG 1738
QY 299 ArgLeuIleAsnAsnAsnTyrPheArgProLysArgLeuAsnPheLysLeuPheAsn 318
DB 1739 CGACTCATCAACAATAACTGGGGATTCGGGCCCAAGAGACTCAACTTCAAGCTCTTCAAC 1798
QY 319 IleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThr 338
DB 1799 ATCCAGGTCAAGGAGGTCAAGAGATGAAGGCAACCAAGACCATCGCAATAACTTACC 1858
QY 339 SerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAla 358
DB 1859 AGCAGGATTCAGGTGTTTACGAGCTCGGAATACAGCTCCGCTACGTCTCGGCTCTCG 1918
QY 359 HisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyr 378
DB 1919 CACCAGGCTCCCTCCCTCCGTCGCGGAGCTCTTCATGATTCTCAGTACGGGTAT 1978
QY 379 LeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerPheTyrCysLeuGluTyr 398
DB 1979 CTGACCTTAAACAATGGCAGTCAGGCTGTGGCCGCTTCTCTCTTCTACTGCTGGATAC 2038
QY 399 PheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGlu 418
DB 2039 TTCCCTCTCAATGCTGAGGAGCGGCAACAACCTTGAATTCAGTACACCTTCGAGGAC 2098
QY 419 ValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeu 438
DB 2099 GTGCTTTTCCACAGCAGCTACGCGCAGCAGCGCTGGACCGGCTGATGAACCTCTC 2158
QY 439 IleAspGlnTyrLeuTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsn 458
DB 2159 ATCCAGCAGTACCTGTATTACTTCCAGAACTCAGTCCACAGGAGGAACTCAAGTACT 2218
QY 459 LysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTyr 478
DB 2219 CAGCAATGTTATTTCTCAAGCGGGCCCGCAACATGTCGGCTCAGGCGCAAGAACTGG 2278
QY 479 LeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 498
DB 2279 CTACCTGAGCGTGTATACCGTCAGCAACAGAGTTTCCACGACACTGTCGCCAACAACAA 2338
QY 499 SerAsnPheThrTyrThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGlySerIleIle 518
DB 2339 AGCAATTTGCTTGGACCGGTGCCACCAAGTATCACCTGAATGGCAGAGACTCCCTGGTT 2398
QY 519 AsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSer 538
DB 2399 AATCCGCGCTTGGCATGGCTACCCACCAAGGACGACGAGGAGCGCTTCTCCCGTCAAGC 2458
QY 539 GlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnVal 558
DB 2459 GGAGTTCTAATGTTTGGCAAGCGGGCTCGAAAGCAATGTGGACTACAGAGCAGTATGGT 2518
QY 559 MetIleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGly 578
DB 2519 ATGCTCACCAGGAAGAAATTAATACTACTAACCCAGTGGCTTACAGAGCAGTATGGT 2578
QY 579 ThrValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAla 598
DB 2579 GTGGTGGCAGACAACCTCGCAGCAGCAACCGAGCTCCCATTTGTGGGAACCTGTCAACAGC 2638
QY 599 MetGlyAlaLeuProGlyMetValTyrGlnAspArgAspValTyrLeuGlnGlyProIle 618
DB 2639 CAGGGGGCTTACCTGGTATGTTCTGGCAAAACCGGAGCTGTACTCGAGGGGCCCATC 2698
QY 619 TrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPhe 638

Db 2699 TGGGCCAAATTCCTCACAGCGACGCACTTTCATCTCTCGCGCTGATGGAGGCTTT 2758
Qy 639 GlyLeuLysAsnProProGlnLeuLeuLeuLysAsnThrProValProAlaAsnPro 658
Db 2759 GGACGTGAACACCCGCCCTCTCAGATCTGGTGAAGAAACACTCTCTGTTCCTCGCGATCCT 2818
Qy 659 ProAlaGluPheSerAlaThrLysPheAlaSerPheLeuThrGlnTyrSerThrGlyGln 678
Db 2819 CGACACCTTCAGCGCCGACGCTGGCTCTTTTATACCGCAGTACAGCACCGGACAG 2878
Qy 679 ValSerValGluLeuLeuGluLeuGlnLysGluAsnSerLysArgTrpAsnProGlu 698
Db 2879 GTGAGCTGGAATCGAATGGAGCTGCAGAAAGAAACAGCAAGCGCTGGAAACCCAGAG 2938
Qy 699 ValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsn 718
Db 2939 ATTGAGTATACTTCAACTACTTACAAATCTACAAATGTGGACTTTGCTGTCATTAATCTG 2998
Qy 719 GlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 2999 GGTACTATTACAGCGCTGCCCATTTGGCACTCGTATCTCACCGGTAATCTG 3052

RESULT 13

US-10-291-583-1
; Sequence 1, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alvira, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
; TITLE OF INVENTION: Sequences and Isolating No. US20030138772A1el Sequences Identifi
; FILE REFERENCE: UPN-02735USA
; CURRENT APPLICATION NUMBER: US/10/291,583
; PRIOR FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 4721
; TYPE: DNA
; ORGANISM: adeno-associated virus serotype 7
US-10-291-583-1

Alignment Scores:
Pred. No.: 0 Length: 4721
Score: 3442.50 Matches: 630
Percent Similarity: 92.02% Conservative: 50
Best Local Similarity: 85.25% Mismatches: 54
Query Match: 86.30% Indels: 5
DB: 13 Gaps: 4

US-09-807-802A-13 (1-736) x US-10-291-583-1 (1-4721)

Qy 1 MetAlaAlaAspGlyTyrLeuProAspTrpLeuGluAspAsnLeuSerGluGlyLeuArg 20
Db 2222 ATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGGACCACTCTCTGAGGGCATTCGC 2281
Qy 21 GluTrpTrpAspLeuLysProGlyAlaProLysProLysAlaAsnGlnLysGlnAsp 40
Db 2282 GAGTGGTGGACCTGAAACCTGGAGCCCGGAAACCCAAAGCCAAACAGCAAGAGCAGGAC 2341
Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
Db 2342 AACGGCGGGGCTGGTCTTCTGGCTACAAAGTACCTCGGACCTTCAACGAGACTCGAC 2401

Qy 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaLysAsp 80
Db 2402 AAGGGGAGAGCCCTCAACGCGCGGACGACGCGCCCTCGAGCAGCAGCAAGGCTACGAC 2461
Qy 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
Db 2462 CAGCAGCTCAAAGCGGTGACAAATCCGTACCTCGGTATTAACACGCGCGAGCTTT 2521
Qy 101 GlnGluArgLeuGlnLysAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln 120
Db 2522 CAGAGCGCTCTGCAAGAAAGATACGTCTTTTGGGGCAACCTCGGGCGAGCAGCTTTCAG 2581
Qy 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGlyAlaLysThrAlaPro 140
Db 2582 GCCAAGAAAGCGGTTCCTCGAACCTCTCGGTCTGGTTCAGGAAGCGCTTAAGACGCGCTCT 2641
Qy 141 GlyLysLysArgProValGluGlnSerProGln---GluProAspSerSerSerGlyLeu 159
Db 2642 GCAAAGAAAGACCGGTAGAGCGCTCACTCAGCGTTTCCCCCGACTCTCTCCACGGGATC 2701
Qy 160 GlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSer 179
Db 2702 GGCAGAAAGCGCAGCAGCGCCCGCAAGAGACTCAATTTCCGTACAGCTGGCGACTCA 2761
Qy 180 GluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGly 199
Db 2762 GAGTCAGTCCCGGACCTCAACCTCTCGGAGAACCTCCAGCAGCGCTCTCTAGTGTGGGA 2821
Qy 200 ProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAsp 219
Db 2822 TCTGTGTCAGTGGTGCAGCGCGGTGGCGACCAATGGCAGCAATAAAGAGGTGGCGAC 2881
Qy 220 GlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgVal 239
Db 2882 GGAGTGGGTATGCTCAGGAATTTGGCATTTCCGATTCGATGGTGGGCGACAGAGTC 2941
Qy 240 IleThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGln 259
Db 2942 ATTACCACGACGACCCGAACTGGGCGCTGCGCCACTTACAAACCAACCACTCTACAGCNA 3001
Qy 260 IleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrPro 279
Db 3002 ATCTCCAGTGAACCTGCAGGTAGTACCAACGACCAACCTACTTCCGGCTACAGCACCCC 3061
Qy 280 TrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArg 299
Db 3062 TGGGGGTATTTTGACTTTTACAGATTCCACTGCCACTTCTCACCACTGACTGGCAGCGA 3121
Qy 300 LeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIle 319
Db 3122 CTCTCAACAACTGCGGATTCGCGCCCAAGAGCTGCGGTTCAGCTCTTCAACATC 3181
Qy 320 GlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSer 339
Db 3182 CAGGTCAAGGAGGTGACGACGAATGACGCGCTTACGACCATCGCTTAATACCTTACCAGC 3241
Qy 340 ThrValGlnValPheSerAspSerGlyTyrGlnLeuProTyrValLeuGlySerAlaHis 359
Db 3242 ACGATTCAAGTATTTCTCGGACTCGGAATACAGCTGCGCTCGCTCTCGGCTCTCGCGAC 3301
Qy 360 GlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeu 379
Db 3302 CAGGCGCTGCTGCTCCGTTCCGCGGACGCTCTTCATGATTCTCTCAGTACGCGCTACCTG 3361
Qy 380 ThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhe 399
Db 3362 ACTCTCAACAAATGGCAGTCAGTCTGTGGAGCGTTCTCTCTTACTGCTGAGGACTTTC 3421
Qy 400 ProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluVal 419
Db 3422 CCCTCTCAGATGCTGAGAACCGGCAACAACTTTTNGTTCAGCTACGCTTCGAGGACGCTG 3481

```
QY 420 ProPheHisSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIle 439
Db 3482 CCTTTCCAGCAGCTACGCACACAGCCGCGGTGATGAATCCCTTCATC 3541
QY 440 AspGlnTyrLeuTyrTyrLeuAsnArgThrGln--AsnGlnSerGlySerAlaGlnAsn 458
Db 3542 GACCAGTACTTGTTACTCTGGCCAGACACAGAGTAACCCAGGAGGCACAGCTGCAAT 3601
QY 459 LysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrp 478
Db 3602 CGGGAATCTCAGTTTACCGGGCGGCGCTTCAACTATGCGCCAAACAGCCAAAGATTGG 3661
QY 479 LeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 498
Db 3662 TTACCTGGACCTTGCTTCGGCAACCAAGAGTCTCCAAACGCTGGATCAAAACAACAAC 3721
QY 499 SerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIle 518
Db 3722 AGCAACTTTGCTTGGACTGTGTGCCACCAATATCATCTGAACGGCAGAAACTCGTTGGTT 3781
QY 519 AsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSer 538
Db 3782 AATCCCGCGCTGCCATGCACTCAAGGAGCAGCAGCGCTTTTTCCTATCCAGC 3841
QY 539 GlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsn--ThrAlaLeuAspAsn 557
Db 3842 GGAGTCTGATTTTGGAAAA-----ACTGGAGCAACTTAACAAACTACATCTGGAAAT 3895
QY 558 ValMetIleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 577
Db 3896 GTGTTAATGACAATGAAGAAGAAATTCGTCTACTAATCTGTAGCCAGGAAGAAATAC 3955
QY 578 GlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHis 597
Db 3956 GGGATAGTCAGCAGCAACTTACAAAGCGGTAACTATGCAGCCAGCAGCAAGTTGTCAAC 4015
QY 598 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro 617
Db 4016 AACGAGGAGCTTACTTGGCATGGTCTGGCAGAACCGGGACGTGTACTTCAGGGTCCC 4075
QY 618 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 637
Db 4076 ATCTGGGCCAAGATTCCTCACAGGATGGCAACTTTCACCCGTCTCTTTGATGGCGGC 4135
QY 638 PheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsn 657
Db 4136 TTTGGACTTAAACATCGCTCTCTAGATCTCTGATCAAGAACACTCCCGTTCCTCGTAAT 4195
QY 658 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 677
Db 4196 CCTCCGAGGTGTTACTCTCGCAAGTTTGTCTTCATCACAGTACAGCACCGGA 4255
QY 678 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPro 697
Db 4256 CAAGTCAGCGTGGAAATCGAGTGGAGCTGCAGAAAGGAAACAGCAAGCGCTGGAAACCG 4315
QY 698 GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsn 717
Db 4316 GAGATTCAGTACACTCTCAACTTGAANAAGCAGACTGGTGTGAGCTTTGCCGTITGACGC 4375
QY 718 AsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 4376 CAGGTTGTTTACTCTGAGCTCGCCCTATTGGCACTGTTTACCTCACCCTGTAATCTG 4432
RESULT 14
US-10-291-583-47
; Sequence 47, Application US/10291583
; Publication No. US20030138772A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Guangping
; APPLICANT: Wilson, James M.
; APPLICANT: Alviria, Mauricio
; TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
```

```
; TITLE OF INVENTION: Sequences and Isolating No. US20030138772A1e1 Sequences Identif
; FILE REFERENCE: UPN-0273SUSA
; CURRENT APPLICATION NUMBER: US/10/291,583
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/350,607
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/341,117
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/377,066
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/386,675
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47
; LENGTH: 3128
; TYPE: DNA
; ORGANISM: 44.5
US-10-291-583-47
```

```
Alignment Scores:
Pred. No.: 0 Length: 3128
Score: 3441.00 Matches: 626
Percent Similarity: 91.06% Conservative: 46
Best Local Similarity: 84.82% Mismatches: 64
Query Match: 86.26% Indels: 2
DB: 13 Gaps: 2
```

US-09-807-802A-13 (1-736) x US-10-291-583-47 (1-3128)

```
QY 1 MetalAlaAspGlyTyrLeuProAspTrpLeuGluAspAsnLeuSerGluGlyIleArg 20
Db 844 ATGGCTCGCATGGTTATCTTCAGATTGGCTCGAGGACCAACTCTCTGAGGCAATTCG 903
QY 21 GluTrpTrpAspLeuLysProGlyAlaProLysProLysAlaAsnGlnLysGlnAsp 40
Db 904 GAGTGGTGGGACTTGAACCTGGAGCCCGCAACCCAAAGCCCAACAGCAAAAGCAGGAC 963
QY 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
Db 964 GAGCGCCGGGGTCTGGTGTCTTCTGGCTTACAGTACCTTCGGACCTTCAACGAGCTCGAC 1023
QY 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
Db 1024 AAGGGGAGCCCGTCAACGCGCGAGCAGCGGCCCTCGAGCAGCAAGCCCTACGAC 1083
QY 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspLysGluPhe 100
Db 1084 CAGCAGCTCAAAGCGGGTGACAATCCGTACCTGCGGTATTAACCCAGCCGAGCTTT 1143
QY 101 GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln 120
Db 1144 CAGGAGCTGTGCAAGAAGATACGTCTTTTGGGGCAACCTTCGGGCGAGCAGTCTTCAG 1203
QY 121 AlaLysLysArgValLeuGluProLeuGlyLeuValGluGlyAlaLysThrAlaPro 140
Db 1204 GCCAAGAAGCGGGTCTCGAACCTCTCGTCTGGTTGAGGAAGCGCTTAAGACGGCTCT 1263
QY 141 GlyLysLysArgProValGluGlnSerProGln---GluProAspSerSerSerGlyIle 159
Db 1264 GGAAAGAGAGACCGGTAGAGCCATCCCGCCAGCGCTTCCACAGCTCTCTACGGGCAATC 1323
QY 160 GlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSer 179
Db 1324 GGCAGAAGGCGCAGCAGCCCGGAAAAGAGACTCAACTTTGGGCGAGACTGCGGACTCA 1383
QY 180 GluSerValProAspProGlnProLeuGlyGluProAlaThrProAlaAlaValGly 199
Db 1384 GAGTCAGTCCCGGACCTCAACCAATCGAGAGAACCCCGCGCAGGCGCCCTCTGGTCTGGA 1443
QY 200 ProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAsp 219
Db 1444 TCTGGTACAATGGCTGCGAGGCGGTGGCGCTCCAAATGGGAGACAATAACAGGCGCCGAC 1503
```

```
Qy 220 GlyValGlyAenAlaSerGlyAenTrpHisCysAaspSerThrTrpLeuGlyAaspArgVal 239
Db 1504 GGAGTGGGTAGTTCCTCAGGAAATGGCATTCGATTCACATGGCTGGGGCAGAGATC 1563
Qy 240 IleThrThrSerThrArgThrTrpAlaLeuProThrTyrAenAenHisLeuTyrLysGln 259
Db 1564 ATCACCAACAGCAGCCCGAACCCTGGGCGCTCCCAACCTACAAACACCACTCTCAAGCAA 1623
Qy 260 IleSerSerAlaSerThrGlyAlaSer---AenAenHisTyrPheGlyTyrSerThr 278
Db 1624 ATCTCCAAACGGGACTTCGGGAGGAAGCACCACGACCAACCTTCTTCGGCTACAGCAC 1683
Qy 279 ProTrpGlyTyrPheAaspPheAenArgPheHisCysHisPheSerProArgAaspTrpGln 298
Db 1684 CCCTGGGGGTATTTGACTTTAAACAGATTCACCTGCCACTTCTCACCGGTGACTGGCAG 1743
Qy 299 ArgLeuIleAenAenAnfTrpGlyPheArgProLysArgLeuAenPheLysLeuPheAen 318
Db 1744 CGACTCATCAACAACACTGGGGATTCGGGCCCAAGAGACCCAACTTCAAGCTCTTCAAC 1803
Qy 319 IleGlnValLysGluValThrAsnAaspGlyValThrThrIleAlaAenAenLeuThr 338
Db 1804 ATCCAGGTCAAGGAGGTCCACGAGATGAAGGCACCAAGACCATCGCAATTAACCTTACC 1863
Qy 339 SerThrValGlnValPheSerAaspSerGluTyrGlnLeuProTyrValLeuGlySerAla 358
Db 1864 AGCAGGATTCAGGCTTTACGGACTCGGAATACAGCTCCGCTACGCTCTCGGCTCTCG 1923
Qy 359 HisGlnGlyCysLeuProPheProAlaAaspValPheMetIleProGlnTyrGlyTyr 378
Db 1924 CACCAGGGCTGCCTGCCTCCGTCCTCCGGCGGACGCTTTCATGATTCCTCAGTAGCGGTAC 1983
Qy 379 LeuThrLeuAenAenGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyr 398
Db 1984 CTGACTCTGAACAATGGCAGTCCAGCCGTGGCGGCTTCCTCTTCTACTGCTCGAGTAC 2043
Qy 399 PheProSerGlnMetLeuArgThrGlyAenAenPheThrPheSerTyrThrPheGluGlu 418
Db 2044 TTTCTCTCTCAATGCTGAGAACCGGGCAACAACCTTTGAGTTTCAGTACCAGTTTGGAGAC 2103
Qy 419 ValProPheHisSerSerTyrAlaHisSerGlnSerLeuAaspArgLeuMetAenProLeu 438
Db 2104 GTGCTTTTTCAGCAGCAGTACGCGCACAGCCAGCCAGCTGGACCGGCTGATGAACCCCTC 2163
Qy 439 IleAaspGlnTyrLeuTyrTyrLeuAenArgThrGlnAenGlnSerGlySerAlaGlnAen 458
Db 2164 ATCGACCAAGTACCTGTACTACTGCTCGGACTCAGTCCACGGGAGGTACCGCAGGAACT 2223
Qy 459 LysAaspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAenTrp 478
Db 2224 CAGCAGTTTGTCTATTTCTCAGCGCGGGCTTAATAACATGTCCGGCTCAGGCCCAAACTGG 2283
Qy 479 LeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAaspAenAenAen 498
Db 2284 CTACCGGGCGCTGCTACCGGCAGCAACCGCTCTCCAGCAGCTGTGCGCAAAATAACAAC 2343
Qy 499 SerAenPheThrTrpThrGlyAlaSerLysTyrAenLeuAenGlyArgGlySerIleIle 518
Db 2344 AGCAACTTTGCTGGACCGGTGCCACCAAGTATCATCTGATGGCAGAGACTCTCTGGTA 2403
Qy 519 AenProGlyThrAlaMetAlaSerHisLysAaspAaspGluAaspLysPhePheProMetSer 538
Db 2404 AATCCCGGTGTGGCTATGTGCAACCCCAAGCAGCAGCAAGACGATTTTTTCCGTCAGC 2463
Qy 539 GlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAenThrAlaLeuAaspAenVal 558
Db 2464 GGAGTCTTTAATGTTTGGGAAACAGGAGCTGGAAAGAACACGCTGGACTATAGCAGCGTT 2523
Qy 559 MetIleThrAaspGluGluGluIleLysAlaThrAenProValAlaThrGluArgPheGly 578
Db 2524 ATGCTAACCACTGAGGAAGAAATTTAAACCAACCAACCCAGTGGCCACGAGACATGACGGC 2583
```

```
Qy 579 ThrValAlaValAenPheGlnSerSerSerThrAaspProAlaThrGlyAaspValHisAla 598
Db 2584 GTGGTGGCGATAAACCTGCAACAGCAAAACCGCGCTCTTATTGTAGGGCGGTCACAGT 2643
Qy 599 MetGlyAlaLeuProGlyMetValTrpGlnAaspArgAaspValTyrLeuGlnGlyProIle 618
Db 2644 CAAGGAGCCTTACTCTGGCATGGTCTGGCAGAACCGGACGCTGACCTGCGAGGGTCTTATC 2703
Qy 619 TrpAlaLysIleProHisThrAaspGlyHisPheHisProSerProLeuMetGlyGlyPhe 638
Db 2704 TGGGCCCAAGATTCTCACACGGACGGAAACTTTTCATCCCTCGCGCTGATGGGAGGCTTT 2763
Qy 639 GlyLeuLysAenProProGlnIleLeuIleLysAenThrProValProAlaAenPro 658
Db 2764 GGACTGAACACACCGCGCTCTCTCAGATCTCTGATTAAGAATACACCTGTTCGCGGATCTCT 2823
Qy 659 ProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGln 678
Db 2824 CCAACTACCTTTCAGTCAAGCTAAGCTGGCGCTGCTTCATCAGCAGTACACACCGGACAG 2883
Qy 679 ValSerValGluIleGluTrpGluLeuGlnLysGluAenSerLysArgTrpAenProGlu 698
Db 2884 GTCAGCGTGAATTTGAATGGGAGCTGCAGAAAGAAACAGCAACGCTGGAACCCACAG 2943
Qy 699 ValGlnTyrThrSerAenTyrAlaLysSerAlaAenValAaspPheThrValAaspAenAen 718
Db 2944 ATTCATATACACTTCCAACTACTACAAATCTACAAATGTGGACTTGTCTGTTAAACACAGAT 3003
Qy 719 GlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 3004 GGCATTATTCAGGCTCCGCCATTTGGCACCGGTACTCCCGCTAATCTG 3057
```

RESULT 15

US-10-291-583-28
Sequence 28 Application US/10291583
Publication No. US2003013872A1

GENERAL INFORMATION:

APPLICANT: Gao, Guangping
APPLICANT: Wilson, James M.
APPLICANT: Alvira, Mauricio
TITLE OF INVENTION: A Method of Detecting and/or Identifying Adeno-Associated Virus
TITLE OF INVENTION: Sequences and Isolating No. US2003013872A1
FILE REFERENCE: US-02735USA
CURRENT APPLICATION NUMBER: US/10/291,583
CURRENT FILING DATE: 2002-11-12
PRIOR APPLICATION NUMBER: US 60/350,607
PRIOR FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 60/341,117
PRIOR FILING DATE: 2001-12-17
PRIOR APPLICATION NUMBER: US 60/377,066
PRIOR FILING DATE: 2002-05-01
PRIOR APPLICATION NUMBER: US 60/386,675
PRIOR FILING DATE: 2002-06-05
NUMBER OF SEQ ID NOS: 120
SOFTWARE: PatentIn version 3.1
SEQ ID NO 28
LENGTH: 3128
TYPE: DNA
ORGANISM: new AAV serotype, clone 42.15
US-10-291-583-28

Alignment Scores:
Pred. No.: 0 Length: 3128
Score: 3440.00 Matches: 626
Percent Similarity: 90.92% Conservative: 45
Best Local Similarity: 84.82% Mismatches: 65
Query Match: 86.24% Indels: 2
DB: 13 Gaps: 2

US-09-807-802A-13 (1-736) x US-10-291-583-28 (1-3128)

Qy 1 MetAlaAaspGlyTyrLeuProAaspTrpLeuGluAaspAenLeuSerGluGlyIleArg 20

Db 844 ATGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGGACAACTCTCTGAGGGCAATTGCG 903
Qy 21 GluTrpTrpAspLeuLysProGlyValaProLysProLysAlaAenGlnGlnLysGlnAsp 40
Db 904 GAGTGGTGGGACTTGAACCTGGAGCCGCCAAACCCAAAGCCAAACAGCAAGCAGGAC 963
Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAenGlyLeuAsp 60
Db 964 GACGGCGGGGTCTGGTCTCTCTGGCTACAGTACCTCGGACCTTCAACGGGACTCGAC 1023
Qy 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
Db 1024 AAGGGGAGCCGCTCAACGGCGGAGCAGCGGCCCTCGAGCAGCAGCAAGGCCCTACGAC 1083
Qy 81 GlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPhe 100
Db 1084 CAGCAGCTCAAAAGCGGTGACAACTCCGTACCTGGGTATTAACCAACCGCCGAGCGAGTTT 1143
Qy 101 GlnGluArgLeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGln 120
Db 1144 CAGGAGCGCTCGCAAGAGATACGCTTTTGGGGGCAACCTCGGGCGGAGCAGTCTTCCAG 1203
Qy 121 AlaLysLysArgProValGluGlnSerProGln---GluProAspSerSerSerGlyIle 159
Db 1264 GGAAGAAGAGAGCCGATGAGCATCACCCAGCGCTTCTCCAGACTCTCTACGGGCATC 1323
Qy 160 GlyLysThrGlyGlnGlnProAlaLysLysArgLeuAenPheGlyGlnThrGlyAspSer 179
Db 1324 GGCAGACAGGCGCAGCAGCCCGGAAAGAGACTCAACTTTTGGGCAGACTGGCGGACTCA 1383
Qy 180 GluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGly 199
Db 1384 GAGTCAGTCCCGACCCCTCAACCAATCGGAGAACCCCGCAGGCCCTCTGGTCTGGGA 1443
Qy 200 ProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAsp 219
Db 1444 TCTGGTACAAATGGCTGACGCGGTGGCGCTCCAAATGGCAGACAATAACGAAGGCCCGAC 1503
Qy 220 GlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTyrLeuGlyAspArgVal 239
Db 1504 GGAGTGGGTAGTTCCTCAGAAATGGCAATTCGGATTCCAGATGCTGGCGCAGAGATC 1563
Qy 240 IleThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGln 259
Db 1564 ATCACCAACAGCAGCCGACCTGGGCCCTCCCACTCAACCAACCACTCTTACAAGCAA 1623
Qy 260 IleSerSerAlaSerThrGlyAlaSer---AsnAspAsnHisTyrPheGlyTyrSerThr 278
Db 1624 ATCTCCAAACGGGACATCGGGGAGGAGCACAACCAACCACTACTCTCGGCTACAGCAC 1683
Qy 279 ProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGln 298
Db 1684 CCTGGGGGTATTTTGACTTTTACAGATTCCACTGCGCACTTCTCCACAGTGCAGTGGCAG 1743
Qy 299 ArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsn 318
Db 1744 CGACTCATCAACAACCTGGGGATTCCGGCCCAAGAGACTCAACTTCAAGCTCTTCAAC 1803
Qy 319 IleGlnValLysGluValThrThrAsnAspGlyValThrThreAlaAsnLeuThr 338
Db 1804 ATCCAGGTCAAGGAGGTCAAGCAATGAAGGACCAACCAACCACTCCCAATACCTTACC 1863
Qy 339 SerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAla 358
Db 1864 AGCACGATTCAGTCTTTACGGACTCGGAATACAGCTCCCGTACGCTCTCGGCTCGG 1923
Qy 359 HisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyr 378
Db 1924 CACCAGGGCTGCCGCTCGGTTCCCGCGGAGCGTCTTATGATTCTCTCAGTACGGGTAC 1983

Qy 379 LeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyr 398
Db 1984 CTGACTCTGAACAAACGGCAGTCAGGCGGTGGCGGTCTCTCTCTACTCTCTGGAGTAC 2043
Qy 399 PheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGlu 418
Db 2044 TTTCTCTCTCAATGCGGAGAACGGGCACAACTTTGAGTTTCAGCTACCAAGTTTGGAGC 2103
Qy 419 ValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeu 438
Db 2104 GTGCTCTTTTCAACGAGCTACGCGCATAGCCAAAGCTGGACCGGTGATGAACCCCTC 2163
Qy 439 IleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsn 458
Db 2164 ATGACACGACTCTGTACTACTCTCTCGACTCTCTCGACTCAGTCCACGGGAGGTACCGCAGAACT 2223
Qy 459 LysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrp 478
Db 2224 CAGCAGTTGCTATTTCTCAGGCGCGGCTTAATAACATGTCTCGCTCAGGCGCAAACTGG 2283
Qy 479 LeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 498
Db 2284 CTACCCGGGCTGTCTACCGGCAGCAACGCGTCTCCACGACACTGCTCGCAAAATAACAAC 2343
Qy 499 SerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIle 518
Db 2344 AGCAACTTTCCTGGACCGGTGCCCACTCATCTGAATGGCAGAGACTCTCTGGTA 2403
Qy 519 AsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSer 538
Db 2404 AATCCCGGTGCTGCTATGGCAACGCAAGCAGCAGCAGGCGATTTTTTCCATCCAGC 2463
Qy 539 GlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnVal 558
Db 2464 GGAGTCTTGATGTTGGNAACAGGGAGCTGGAAAGACCAACCTGGACTATAGCAGCGTT 2523
Qy 559 MetIleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGly 578
Db 2524 ATGCTAACAGTGGAGGAATCAAAACCAACCCAGTGGCCACAGACAGTACGCGC 2583
Qy 579 ThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAla 598
Db 2584 GTGGTGGCGGATACTGCAACAGCAAAACGCCCTCTCTATTGTAGGGCGCGTCAACAGT 2643
Qy 599 MetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIle 618
Db 2644 CAAGGAGCCTTACCTGGCATGGTCTGGCAGAACCGGACGCTGACCTCGAGGGTCTATC 2703
Qy 619 TrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPhe 638
Db 2704 TGGGCAAGATCTCTCACCGGACGGCACTTTTCATCTCTCGCGCTGATGGAGGCTTT 2763
Qy 639 GlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsnPro 658
Db 2764 GGACTGAACACCGGCTCTCTCAGATCTGATTAAGAATAACACCTGTTTCCCGGATCCT 2823
Qy 659 ProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGln 678
Db 2824 CCAACTACTCTCAGTCAAGCCCAAGCTGGCTGCTCATCAGCAGTACAGCACCGGACAG 2883
Qy 679 ValSerValGluIleGluTrpGluGlnLysGluAsnSerLysArgTrpAsnProGlu 698
Db 2884 GTCAGCTGGAAATTTGAATGGAGCTGCAGAAAGAGACAGCAAGCGCTGGAAACCCAGAG 2943
Qy 699 ValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsn 718
Db 2944 ATTCAGTATACTTCCAACTACTACAAATCTCAAAATGTGGACTTTTGTGTCTAATACTAG 3003
Qy 719 GlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 736
Db 3004 GGTACTTATTCAGAGCCTCGCCCCCATTTGGACCCGCTTACCTCACCCTGTAACCTG 3057

Sun Feb 15 18:36:55 2004

us-09-807-802a-13.rnpb

Page 21

Search completed: February 15, 2004, 02:04:02
Job time : 852.807 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 14, 2004, 15:26:50 ; Search time 4084.43 Seconds
(without alignments)
4379.586 Million cell updates/sec

Title: US-09-807-802A-13
Perfect score: 3989
Sequence: 1 MAADGYLPDLWLDNLSEGR.....NNGLYTEPRPIGTRYLTRPL 736

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09807802/runat_11022004_175608_15941/app_query.fasta_1.2389
-DB=EST -QFMT=FASTAP -SUPPLX=rest -MINMATCH=0.1 -LOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=spt -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=spt -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09807802.cgn 1 1 7257 @runat_11022004_175608_15941 -NCPU=3
-NO_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estinv:*
4: em_estin:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID | Description |
|------------|-------|---------------|--------|----|----------|--------------------|
| 1 | 193.5 | 4.9 | 753 | 28 | BH115587 | BH115587 RPCI-24-3 |
| 2 | 147 | 3.7 | 2318 | 11 | BC014681 | BC014681 Homo sapi |
| 3 | 139.5 | 3.5 | 803 | 29 | BZ265284 | BZ265284 CH230-375 |
| 4 | 134 | 3.4 | 856 | 28 | BH164736 | BH164736 ENTTS28TF |
| 5 | 133 | 3.3 | 2146 | 11 | AK042727 | AK042727 Mus muscu |
| 6 | 133 | 3.3 | 2295 | 11 | AK035953 | AK035953 Mus muscu |
| 7 | 129 | 3.2 | 3943 | 11 | AK087510 | AK087510 Mus muscu |
| 8 | 125 | 3.1 | 640 | 14 | CA356072 | CA356072 628112 NC |
| 9 | 125 | 3.1 | 701 | 14 | CA363789 | CA363789 638563 NC |
| 10 | 125 | 3.1 | 745 | 12 | BJ139005 | BJ139005 BJI39005 |
| 11 | 125 | 3.1 | 754 | 10 | BG207226 | BG207226 RST26694 |
| 12 | 125 | 3.1 | 2019 | 11 | AK014322 | AK014322 Mus muscu |
| 13 | 124 | 3.1 | 925 | 13 | BQ440075 | BQ440075 AGENCOURT |
| 14 | 123.5 | 3.1 | 645 | 12 | BI911451 | BI911451 603063370 |
| 15 | 123.5 | 3.1 | 3512 | 11 | AK038988 | AK038988 Mus muscu |
| 16 | 123.5 | 3.1 | 4763 | 11 | AK083225 | AK083225 Mus muscu |
| 17 | 123.5 | 3.1 | 756 | 28 | AQ742711 | AQ742711 HS 5386 B |
| 18 | 122.5 | 3.1 | 925 | 28 | BH152154 | BH152154 ENTVP12TR |
| 19 | 121.5 | 3.0 | 1418 | 13 | BQ686477 | BQ686477 AGENCOURT |
| 20 | 121.5 | 3.0 | 3689 | 11 | AK031988 | AK031988 Mus muscu |
| 21 | 121.5 | 3.0 | 567 | 14 | CA408136 | CA408136 p129c03.y |
| 22 | 120.5 | 3.0 | 629 | 14 | CA408136 | CA408136 pk09b07.x |
| 23 | 120.5 | 3.0 | 629 | 14 | CNS03388 | AL225809 Tetradon |
| 24 | 120.5 | 3.0 | 2080 | 11 | BC001541 | BC001541 Homo sapi |
| 25 | 120.5 | 3.0 | 2080 | 11 | BC001541 | BM000978 BQGU282TF |
| 26 | 120.5 | 3.0 | 812 | 28 | BH600978 | BM000978 AGENCOURT |
| 27 | 120 | 3.0 | 937 | 12 | BM802157 | AK048546 Mus muscu |
| 28 | 120 | 3.0 | 4782 | 11 | AK048546 | AK033012 Mus muscu |
| 29 | 119.5 | 3.0 | 3329 | 11 | AK033012 | BX382214 BX382214 |
| 30 | 119 | 3.0 | 919 | 13 | BX382214 | AG072149 Pan trogl |
| 31 | 118.5 | 3.0 | 973 | 29 | AG072149 | BQ061577 AGENCOURT |
| 32 | 118.5 | 3.0 | 1035 | 12 | BQ061577 | EC032608 Homo sapi |
| 33 | 118 | 3.0 | 3344 | 11 | BC032608 | AK076994 Mus muscu |
| 34 | 117 | 2.9 | 5809 | 11 | AK076994 | BQ645964 AGENCOURT |
| 35 | 116.5 | 2.9 | 1085 | 13 | BQ645964 | AK049740 Mus muscu |
| 36 | 116.5 | 2.9 | 2249 | 11 | AK049740 | BC030211 Homo sapi |
| 37 | 115 | 2.9 | 1604 | 11 | BC030211 | BQ642977 AGENCOURT |
| 38 | 115.5 | 2.9 | 1042 | 13 | BQ642977 | BQ073024 AGENCOURT |
| 39 | 115.5 | 2.9 | 1085 | 13 | BQ716249 | BQ716249 AGENCOURT |
| 40 | 115 | 2.9 | 1133 | 13 | BQ716249 | AK019448 Mus muscu |
| 41 | 115 | 2.9 | 2931 | 11 | AK019448 | AK041115 Mus muscu |
| 42 | 115 | 2.9 | 4784 | 11 | AK041115 | BF984198 602307620 |
| 43 | 114.5 | 2.9 | 641 | 10 | BF984198 | BI838324 603083262 |
| 44 | 114.5 | 2.9 | 936 | 12 | BI838324 | BUS00074 AGENCOURT |
| 45 | 114.5 | 2.9 | 1023 | 13 | BUS00074 | |

ALIGNMENTS

RESULT 1
BH115587
LOCUS RPCI-24-358F16.TV RPCI-24 Mus musculus genomic clone RPCI-24-358F16
DEFINITION RPCI-24-358F16.TV RPCI-24 Mus musculus genomic clone RPCI-24-358F16
ACCESSION BH115587
VERSION BH115587.1 GI:14954954
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 753)

AUTHORS Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M., Teegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E., Russell, D., de Jong, P. and Fraser, C.M.

TITLE Mouse BAC end Sequences from Library RPCI-24

JOURNAL Unpublished

COMMENT Other GSSs: RPCI-24-358F16.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.html>). BAC end plate: 358 row: F column: 16
Seq primer: T7
Class: BAC ends.

FEATURES

source
1. 753
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-358F16"
/sex="Male"
/cell_type="Spleen/Brain"
/clone_lib="RPCI-24"
/note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."
BASE COUNT 211 a 178 c 167 g 197 t
ORIGIN
Alignment Scores:
Pred. No.: 9,71e-07 Length: 753
Score: 193.50 Matches: 68
Percent Similarity: 46.74% Conservativity: 18
Best Local Similarity: 36.96% Mismatches: 64
Query Match: 4.85% Indels: 36
DB: 28 Gaps: 5
US-09-807-802a-13 (1-736) x BH115587 (1-753)

QY 91 LeuArgTyrAsnHisAlaSepAlaGluPheGlnGlu-ArgLeuGlnGluAspThrSerPh 110
Db 212 CTCCTAATAATACACAGAGAGAGAGTTTCAGGAGAAATCCAGTAGTACATCTTTT 271
QY 110 eGlyGlyAsnLeuGlyArgAlaValPheGlnAlaLysLysArgValLeuGluProLeuGI 130
Db 272 T-----TTGGCAGGAACCGTGCCAAATGCCAAAGAAAAGGGTTCTCAAAACCCATCGG 322
QY 130 YLeuValGluGluGlyAlaLysThrAlaProGlyLysLys-----ArgProValGluGI 148
Db 323 CTTGGTCTAAGAAGGATTTCAGCTGGCTCTGGGAAGAAATAGATTTCGACCTTA-AATCT 381
QY 148 nSerPro-----GlnGluProAspSerSerGlytI 159
Db 382 CTCCTCCCATGAGGAGTACACTCGCTCTTTTCAGAGTTATGCCAAAAACAGTAGAAG 441
QY 159 eGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSe 179
Db 442 TGGAGAGGCTGGCTTCTAGTCTATGACACAGATGATGAGTTCACGACAGATATCTTC 501
QY 179 rGluSerValProAspProGlnProLeuGlyGluProAlaThrProAlaAlaValGI 199
Db 502 CAATTTA----- 508

QY 199 YProThrThrMetAlaSerGly---GlyGlyAlaProMetAlaAspAsnAsnGluGlyAl 218
Db 509 -GACACTTCTATAATCTCTGGAGCTGGAGTCTACTAATGGCAACAACACAGAC-AC 566
QY 218 aSepGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspAr 238
Db 567 TGATGGAGTGAGCAATGCCATGGATGATGGCATTCGCATTCCTCAAGTGTATGGTAGACTG 626
QY 238 gValileThrThrSerThrArgThrTrpAlaLeuProThrThrAsnAsnHisLeuTyrLy 258
Db 627 AGTCATTACCTAATCTCCAGAGACCTGGTCTGCTGCCCACTCCCACTTAACATTCCACAA 686
QY 258 sGlnIleSer 261
Db 687 CTATATGAAC 696

RESULT 2

BC014681
LOCUS BC014681
DEFINITION Homo sapiens, Similar to oea, clone IMAGE:3866722, mRNA.
ACCESSION BC014681
VERSION BC014681.1 GI:15928395
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2318)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (01-OCT-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villalona@bcm.tmc.edu.
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

FEATURES

source
1. 2318
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3866722"
/tissue_type="Eye, retinoblastoma"
/clone_lib="NIH MGC_67"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
BASE COUNT 637 a 680 c 544 g 457 t
ORIGIN
Alignment Scores:
Pred. No.: 0.0559 Length: 2318
Score: 147.00 Matches: 172
Percent Similarity: 32.27% Conservativity: 91
Best Local Similarity: 21.10% Mismatches: 273
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
Series: IRAP Plate: 21 Row: 1 Column: 9
This clone was selected for full length sequencing because it passed the following selection criteria: Genomescan gene prediction
This clone has the following problem: frame shifted.


```

Db 1970 CACCTCCACACCTCATACACAGCTTGGGACAGCTCCACCATCACCC----- 2019
Qy 637 lypheGlyLeuLysAsnProProGlnleLeuLeuLysAsnThrProValProAlaA 657
Db 2020 -----CAGAACACCCCGCAGAAAGTG-----CCTGTGCATCAGC 2053
Qy 657 snProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrG 677
Db 2054 ATTCCCGTCGAGGCCCTTTCTAGAGAACCCAGTCCGGATAGTACTCAGGTTAGTGAC 2113
Qy 677 lylGlnValSerVal-----GluIleGluTrpGluLeuGlnL 689
Db 2114 CGAATGCTCAGCTAGTGAAGAGTGATTAACCTGCCATCAATAGACAGCCACAC 2173
Qy 689 ysGluAsnSerLysArgTrpAsnProGluValGlnTyrThrSerAsnTyrAlaLysSerA 709
Db 2174 AAAAGAAGAAGAAAAGAAAACCAACCATTTGTAGACAGAGATCCCAAGGTTTGT 2233
Qy 709 laAsnValAspPheThr-----ValAspAsnAsnGlyLeu 720
Db 2234 GTAAAGATGACTTCCTCGTGGGTAGATAACCAAGAACTA 2274

```

RESULT 3

```

BZ265284 803 bp DNA linear GSS 15-OCT-2002
LOCUS CH230-375A22.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
DEFINITION CH230-375A22, genomic survey sequence.

```

```

ACCESSION BZ265284
VERSION BZ265284.1 GI:23978528

```

```

KEYWORDS GSS.

```

```

SOURCE Rattus norvegicus (Norway rat)

```

```

ORGANISM Rattus norvegicus

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

REFERENCE 1 (bases 1 to 803)

```

AUTHORS Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn
A., Gebregeorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de
Jong, P., and Fraser, C.M.

```

```

Rat BAC End Sequences from Library CHORI-230 MbOI segment

```

```

Unpublished

```

```

Other GSSs: CH230-375A22.TV

```

```

Contact: Shaying Zhao

```

```

Department of Eukaryotic Genomics

```

```

The Institute for Genomic Research

```

```

9712 Medical Center Dr., Rockville, MD 20850, USA

```

```

Tel: 301 838 0200

```

```

Fax: 301 838 0208

```

```

Email: szhao@tigr.org

```

```

Clones are derived from the rat BAC library CHORI-230

```

```

(http://www.chori.org/bacpac/rat230.htm). For BAC library

```

```

availability, please contact Pieter de Jong (pjejong@mail.cho.org).

```

```

Clones may be purchased from BACPAC Resources

```

```

(http://www.chori.org/bacpac/or ering information.htm). BAC end

```

```

plate: 375 row: A column: 22

```

```

Seq primer: SP6

```

```

Class: BAC ends.

```

```

FEATURES Location/Qualifiers

```

source

```

1..803

```

```

/organism="Rattus norvegicus"

```

```

/mol_type="genomic DNA"

```

```

/strain="BN/SeNHsd/MCW"

```

```

/db_xref="taxon:10116"

```

```

/clone="CH230-375A22"

```

```

/sex="Female"

```

```

/cell_type="Brain"

```

```

/clone_lib="CHORI-230 Segment 2"

```

```

/notes="Vector: PTARBAC1.3; Site 1: MbOI; Site 2: MboI;

```

```

CHORI-230 Rat (BN/SeNHsd/MCW) BAC library produced by

```

```

Pieter de Jong"

```

```

BASE COUNT 249 a 177 c 178 g 199 t

```

ORIGIN

```

Alignment Scores: 0.0593 Length: 803
Pred. No.: 139.50 Matches: 68
Score: 42.98% Conservative: 33
Percent Similarity: 28.94% Mismatches: 76
Best Local Similarity: 3.50% Indels: 60
Query Match: 29 Gaps: 10
DB:

US-09-807-802A-13 (1-736) x BZ265284 (1-803)
Qy 3 AlaAspGlyTyrLeuProAspTrpLeu-----GluAspAsnLeuSerGluGlyIleArg 20
Db 72 TCTGAGGGAACCTGGCAAGACTGGTGTAGCAAAAGCTTAACCTGAGC-----TTAAAG 125
Qy 21 GluTrpTrpAspLeuLysProGlyAlaProLysPProLysAlaAsnGlnGlnLysGlnAsp 40
Db 126 CAG-----GATAGCAAAAAACACTACCAACCTTAACCCCAATTCTATATCAGGAAACCAAGTC 179
Qy 41 AspGlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAsp 60
Db 180 AATATATCTGGTCTGCTCTTAGGTGC-TATAAGTACTTGGGCTGCCAAAC----- 229
Qy 61 LysGlyGluProValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAsp 80
Db 230 -----CAGTGATTTAATTCAGCAGAC----- 250
Qy 81 GlnGlnLeuLysAlaGlyAspAsnPro-----TyrLeuArgTyr 93
Db 251 ---ATGGTGAAGATCAAGACACACCATCACTCCTCAAGCATAGACTCATATCAAGTAC 307
Qy 94 AsnHisAlaAspAlaGluPheGlnGluArgLeuGlnGluAspThrSerPheGlyGlyAsn 113
Db 308 AGCCTGTGCAAGCAAGAAATTCAGAGAAAATCCAAACAGATGCTTTTCCAGAGGCAT 367
Qy 114 LeuGlyArgAlaValPheGlnAlaLysValArgValLeuGluProLeuGlyLeuValGlu 133
Db 368 CTCAACCAAGTACTCTTCCAGGCTAAAAGAGAGACTTCTGGATCCTTTGACCTGGTTAA- 426
Qy 134 GluGlyAlaLysThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluPro 153
Db 427 AAAGGTTGT----- 438
Qy 154 AspSerSerSerGlyIleGlyLysThrGlyGlnGlnProAlaLysLys-----ArgLeu 171
Db 439 GAGATGTCTCTGGCAGCCAGCAATGACTCTTTTACTCTTCAGCAAAAGCTTCACAGAACT 498
Qy 172 AsnPheGlyGlnThrGlyAspSerGluSerValProAspProGlnProLeuGlyGlu--- 190
Db 499 AATCACTGAAGATGATAAGGTCACTGAACATATATTAAACCAACCTGGATCCCGTCATCT 558
Qy 191 -----ProProAlaThrProAlaAlaValGlyProThrThrMet 203
Db 559 GGTTATGGTGTCAACAAATAGCAACTGCAAAACCCAGCAAACTTGGGGTCTCTTACAAATG 618
Qy 204 Ala-----SerGlyGlyGlyAlaProMetAlaAspAsnAsnGlu 216
Db 619 TCTCTTACACCAGAGGAGGTGTTCCACAAATGCAATAACACAG 663

```

RESULT 4

```

BHI64736
LOCUS

```

```

DEFINITION

```

```

BHI64736

```

```

ACCESSION

```

```

VERSION

```

```

KEYWORDS

```

```

SOURCE

```

```

ORGANISM

```

```

Entamoeba histolytica

```

```

Eukaryota; Entamoebidae; Entamoeba.

```

```

REFERENCE

```

```

1 (bases 1 to 856)
Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.

```

```

BHI64736 856 bp DNA linear GSS 24-SEP-2001
ENTTS28TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, genomic survey sequence.

```

```

ACCESSION BHI64736

```

```

VERSION BHI64736.1 GI:15738174

```

```

KEYWORDS GSS.

```

```

SOURCE Entamoeba histolytica

```

```

Entamoeba histolytica

```

```

Eukaryota; Entamoebidae; Entamoeba.

```

```

REFERENCE 1 (bases 1 to 856)
Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.

```

TITLE Determination of clone end sequences from Entamoeba histolytica
 JOURNAL HMI:IMSS sheared DNA library (2001)
 COMMENT Unpublished
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjlloftus@igr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library

Seq primer: M13-Forward

Class: shotgun

High quality sequence start: 17

High quality sequence stop: 629.

FEATURES

source

1..856

/organism="Entamoeba histolytica"

/mol_type="genomic DNA"

/strain="HMI:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: pHOS1; Site_1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Borell, Oxford University Press, 1999)."

BASE COUNT

347 a 229 c 89 g 191 t

Alignment Scores:

Pred. No.: 0.197 Length: 856
 Score: 134.00 Matches: 88
 Percent Similarity: 33.42% Conservative: 44
 Best Local Similarity: 22.28% Mismatches: 140
 Query Match: 3.36% Indels: 123
 DB: 28 Gaps: 19

US-09-807-802A-13 (1-736) x BH164736 (1-856)

Qy 297 TTPGlnArgLeuIleAsnAsnAsnThrGlyPheArgProLysArgLeuAsnPheLysLeu 316
 |||||
 Db 9 TGGAGACACCTCATTTGGTAAAT-----TCCTTCATTCGTGA 47
 Qy 317 PheAsnIleGlnValLysGluValThrThrAsnAspGlyValThr----- 332
 |||||
 Db 48 TTTAATATAACCAACGTCGGGAACAACTAATCCGTTCAATTCAGTGAATTCAGTC 107
 Qy 333 ---IleAlaAsnLeuThrSerThrValGlnValPheSerAspSerSerGluThrGlnLeu 351
 |||||
 Db 108 TCAGGAGCAATATACCTTTTCTACTACTCAA-----AACATACACAACT 155
 Qy 352 ProTyrValLeuGlySerAlaHisGlnGlyCysLeuProProPheProAlaAspValPhe 371
 |||||
 Db 156 CCTTTCAACAACGAACAACTAAC-----CCATTTAAT----- 188
 Qy 372 MetIleProGlnTyrGlyTyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSer 391
 |||||
 Db 189 -----ACTACTAACAATACTACTAACTCA----- 212
 Qy 392 SerPheTyrCysLeuGluTyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThr 411
 |||||
 Db 213 -----AACAACTAATCACTTTAAT 230

Qy 412 PheSerTyrThrPheGluGluValProPheHisSerSerTyrAlaHisSerGlnSerLeu 431
 |||||
 Db 231 ACTAATACTACAGCACCAACAATCCATTTTAATACAATACTACTAAT----- 278
 Qy 432 AspArgLeuMetAsnProLeuIleAspGlnTyrTyrLeuAsnArgThrGlnAsn 451
 |||||
 Db 279 -----AATCCATTT-----AGTACTAATAATAC 299
 Qy 452 GlnSerGlySerAlaGlnAsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMet 471
 |||||
 Db 300 ACTACCAATGGACAACCAAT-----TTATTTAATCAA-----ACTCCTCTCTCTAAT 347
 Qy 472 SerValGlnProLysAsnThrLeuProGlyProCysTyrArgGlnGlnArgValSerLys 491
 |||||
 Db 348 ACTATAACACACAGGAAT-----AATACAACACTACT 377
 Qy 492 ThrLysThrAspAsnAsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeu 511
 |||||
 Db 378 ACAACAGGAATGATCCATTTGGTAAATTCACCTACCTAAT----- 419
 Qy 512 AsnGlyArgGluSerIleIleAsnProGlyThrAlaMetAlaSerHisLysAspGlu 531
 |||||
 Db 420 AATACTACCTCCAGTCAACTTCAACAGGAACAACCATCAACAGGAAGTAATCCATTT 479
 Qy 532 AspLysPhePheProMetSerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSer 551
 |||||
 Db 480 GGTAAATTTCACTCCCTCAAACTCAGCTCCAAACACAGCAAAATATATACACAGGAACAGCT 539
 Qy 552 AsnThrAlaLeuAspAsnValMetIleThrAspGluGluGluIleLysAlaThrAsnPro 571
 |||||
 Db 540 ACAACACAGGAATAATACACCATCACA-----ACAGGAGTAAATCCA 584
 Qy 572 ValAlaThrGluArgPheGlyThrValAlaValAsnPheGlnSerSerThrAspPro 591
 |||||
 Db 585 -----TTTGGT-----AATTTCACTCCCTCAACACAACTCCC 617
 Qy 592 AlaThrGlyAspValHisAlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAsp 611
 |||||
 Db 618 ACCACAGGAAT-----AATATCAACAGGAACAACCTACCAACAGGAAT 662
 Qy 612 ValTyrLeuGlnGlyProIleTrpAlaLysIleProHisThrAspGlyHisPheHisPro 631
 |||||
 Db 663 AATATACAGGA-----ACAGGAAGTATCATTT-----GGTAATTTCACTACT 707
 Qy 632 SerProLeuMetGlyGlyPheGlyLeuLysAsnProProProGlnIleLeuIle-LysAs 651
 |||||
 Db 708 TCAAACTCACTACCCACACAGGAACAACAACCTACACAGGAATAACACACTACAGGA 767
 Qy 651 nThrProValProAlaAsnProProAlaGluPheSerAlaThrLysPheAlaSerPheIle 671
 |||||
 Db 768 CAACCCCAACACCGGAACACCCCAACCGGAACACCCCAAC----- 810
 Qy 671 eThrGlnTyrSerThrGlyGlnValSerValGluIleGluTrp 685
 |||||
 Db 811 -----ACAGGAGGATCCTTTTGGAAATTTTCATAC 840

RESULT 5

AK042727

LOCUS

DEFINITION

AK042727

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

2146 bp mRNA linear HTC 05-DEC-2002
 Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length
 enriched library, clone:A730019105 product:hypothetical protein,
 full insert sequence.

AK042727

GI:26335284

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

1

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.


```
Qy 135 GlyAlaLysThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAsp 154
Db 723 -----CCCGTCCCGCAGGTGCCCGAGGCGGT 752
Qy 155 SerSerGlyIleGlyLysThrGlyGlnGlnProAlaLysLys----- 169
Db 753 GGTGGCGTCCAGTGGAGT-CTATCCCGGATCCAGAACGAAAGGACGCGCAATCTTT 811
Qy 170 -----ArgLeuAsnPheGlyGlnThrGlyAspSerGluSerValProAspProGln 186
Db 812 TGTGATGGCGCTCCATGTGGGCCCCCGGCGCTTCAGTACCAACAGTACGAGC 871
Qy 187 ProLeuGlyGluProProAlaThrProAlaAlaValGlyProThrMetAlaSerGly 206
Db 872 CCCCTA-----CCAGCAGCAGCAGCGCGCGC---TCCTCGGGCCCCCGCG 919
Qy 207 GlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGlyAsnAlaSerGly 226
Db 920 ACAGGGCCACCCCGACGA-----CATGCAGCAGAT-----GGCAGCTATCTGGC 964
Qy 227 AsnTrpHis-----CysAspSerThrTrpLeuGlyAspArgValIleThrThrSer 243
Db 965 AGTGGGG-ATTCTCCATGCAGCAGCAG-----GCCAGCCAC 1002
Qy 244 ThrArgThrTrpAlaLeuProThrTyraenAsnHisLeuTyrlsGlnIleSerSerAla 263
Db 1003 AGCAGAGTGGC-----CAGTTTCCCAAGGC 1031
Qy 264 SerThrGlyAlaSerAsnAspAsnHisTyrlsPheGlyTyrlsSerThrProTrpGlyTyrlsPhe 283
Db 1032 CAAGAGGGCTCAGTCAGGAAGCCCTTTTCATTGCCACCTCAGGACCC---GGCCACCTG 1088
Qy 284 AspPhe-----Asn 286
Db 1089 TGGCACATGCCCGCAGAGCCCCAGCATGGCCCTTCCTCGGCCACCCAGTGCAGCAG 1148
Qy 287 ArgPheHisCysHis-----Phe 292
Db 1149 CAGTTCCACCATCACCCGCTGCTCTCCAGCGAGAATCGTTGCCACAGTCCCGAGATTC 1208
Qy 293 SerProArgAspTrpGlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeu 312
Db 1209 TCCCCCAACCTCTCTCAAA-----GGGGCTGTGAGGCCCAAAACCTT 1253
Qy 313 AsnPheLysLeuPheAsnIleGlnValLysGluValThrThrAsnAspGlyValThrThr 332
Db 1254 AACTTTAGTTCTCGGAACAGACAGTCCCTCCCTACTGTATAAACAC----- 1301
Qy 333 IleAlaAsnAsnLeuThrSerThrValGlnValPheSerAspSerSerGluTyrlsGlnLeuPro 352
Db 1302 -----TCAGGCGAGTATTCTCGATATCT 1325
Qy 353 TyrValLeuGlySerAlaHisGlnGlyCysLeu----- 363
Db 1326 TAC-----AGTAACCTAAATCAGGATTTAGTTAAACAGTACAGGATGAATCAAAATTTA 1379
Qy 364 -----ProPheProAlaAspValPheMetIleProGlnTyrlsGlyTyrls 378
Db 1380 GGCCTTACAAACAGTACTCTCAATGAATCAGTCGGTA-----CCAAGATAC----- 1424
Qy 379 LeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrlsCysLeuGluTyrls 398
Db 1425 -----CCCAATGCCGTGGGA----- 1439
Qy 399 PheProSerGlnMetLeuArgThrGlyAsnAsnAsnPheThrPheSerTyrlsThrPheGluGlu 418
Db 1440 TTTCCCGTCA-----AACAGTGGTTCAGGGACTCGTG-----CACCAG 1475
Qy 419 ValProGlyHisSerSerTyrlsAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeu 438
Db 1476 GAGCCTATCCATCCAGTGGCTCACTGAACCAATATGACACAACTATGCACCT--- 1532
```

```
Qy 439 IleAspGlnTyrlsLeuTyrlsLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsn 458
Db 1533 -----TCACAGCCTCAGGGAACGTTACGCTCTCCACCTCCC 1568
Qy 459 LysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrp 478
Db 1569 ATGTACCCCATGAAGCAATAGTAACCCAGCAGC-----ACGCCGCTCTCCGACGGTC 1622
Qy 479 LeuProGlyProCysTyrlsArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsn 498
Db 1623 AGGCCCGAAG-TCGTGG-----GAT 1642
Qy 499 SerAsnPheThrTrpThrGlyAlaSerLysTyrlsAsnLeuAsnGlyArgGluSerIleIle 518
Db 1643 GCCCATGGAAGTTGGCAG---TTATCCAAATAT-----GCCCCACCTCAGCCATC 1690
Qy 519 AsnProGlyThrAlaMetAlaSerHisLysAspGluAspLysPhePheProMetSer 538
Db 1691 TCACCCAGCCCCCTGTCCTCCATGGGAATCGGACAGAGGA----- 1729
Qy 539 GlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnVal 558
Db 1730 -----TATGGCCCCCAGAAACATGCAGCAGCCTCGTTGTTTATGGCGCATGTC 1777
Qy 559 MetIleThrAsp-----GluGluGluIleLys----- 567
Db 1778 CTCGCCACCCAGGAGCTGACCGCACACATGAGACCAATGGTTGCCCTGTGTGCGCCT 1837
Qy 568 -----AlaThrAsnProValAlaThrGluArgPheGlyThrValAlaValAsnPhe 584
Db 1838 TGGGATCCACAGCAATCCAGGAACGATGATACCTGGCCCAACGACCC----- 1888
Qy 585 GlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeuProGly 604
Db 1889 TGGTCAGCAGCCATCTTTCCAGCAGTTGCCAACCTGCTCTCCACTACAGCCCCACCC--- 1945
Qy 605 MetValTrpGlnAspArgAspValTyrlsGlnGlyProIleTrpAlaLysIle-ProHi 624
Db 1946 -----AGGCCTGCACCAAGTCTTTCACCTCCACATCCCA 1978
Qy 624 sThrAsp-----GlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAs 642
Db 1979 TCACCCAGCCTTGGGCACAACTCCACCCGTACCC-----CAGAA 2017
Qy 642 nProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPh 662
Db 2018 TACCCCAACAGAAAGTG-----CCTGTGCATCAGCATTTCTCCATCGGAGCC 2062
Qy 662 eSerAlaThrLysPheAlaSerPheIleThrGlnTyrlsSerThrGlyGlnValSerVal-- 681
Db 2063 CTTTCTAGAGAAGCCAGTGGCGGATATGACCCAGGTCCAGCGCACAGAACGCTCAGCTAGT 2122
Qy 682 -----GluIleGluTrpGluLeuGlnLysGluAsnSerLysAr 694
Db 2123 GAACAGCGCAGATTACTCTCGCTCGATAGACAGCCACGAGAGAGAGAGAGAGAAA 2182
Qy 694 gTrpAsnProGluVal 699
Db 2183 GAAAAACAAACCACTT 2198
```

RESULT 7

AK087510

LOCUS

DEFINITION

AK087510.1

GI:26104328

KEYWORDS

HTC; CAP trapper.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

3943 bp mRNA linear HTC 05-DEC-2002
Mus musculus 0 day neonate eyeball cDNA, RIKEN full-length enriched
library, clone: E130308C19 product: weakly similar to TESTIS-SPECIFIC
Y-ENCODED-LIKE PROTEIN (FRAGMENT) [Homo sapiens], full insert
sequence.

ACCESSION

AK087510

VERSION

AK087510.1

KEYWORDS

HTC; CAP trapper.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
9279253
10349636

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, S., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Giesi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G.,
Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M.,
Wagner, J., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Brownstein, M. J., Bult, C.,
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustinich, S., Hill, D.,
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyo-oka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohtsuki, S.
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851

5 The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 3943)

Adachi, J., Aizawa, K., Akimura, T., Hara, A., Hashizume, W.,
Fukuda, S., Furuno, M., Hanagaki, T., Hiraoka, T., Hirozane, T.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, W.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of

Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.

FEATURES
source
1..3943
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:E130308C19"
/db_xref="taxon:10090"
/clone="E130308C19"
/tissue_type="eyeball"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
misc_feature
1..3943
note="weakly similar to TESTIS-SPECIFIC Y-ENCODED-LIKE
PROTEIN (FRAGMENT) [Homo sapiens] (SPT1075885, evidence:
FASTA, 68.8%ID, 65.7%length, match=143)"
BASE COUNT 1004 a 904 c 956 g 1079 t
ORIGIN
Alignment Scores:
Pred. No.: 4.44 Length: 3943
Score: 129.00 Matches: 112
Percent Similarity: 31.57% Conservative: 49
Best Local Similarity: 21.96% Mismatches: 199
Query Match: 3.23% Indels: 150
Gaps: 28
US-09-807-802A-13 (1-736) x AK087510 (1-3943)
QY 22 TtpTppAppLeuLysProGlyAlaProLysProLysAlaAsnGlnGlnLysAlaAsp 41
Db 119 TGGCAGCAGCAAGAGCCCA-----CCACAGAGCCCGCGCTCGGGAGGACTCC 166
QY 42 GlyArgGlyLeuValLeuProGlyTyrLysTyrLeuGlyProPheAsnGlyLeuAspLys 61
Db 167 GCGGCGCGCAGGTGTCAGGCGCGCGCAGCTCAAGGAGGC----- 205
QY 62 GlyGluProValAsnAlaAlaAspAlaAla-----LeuGluHisAspLysAlaTyr 79
Db 206 GCCGAACAGCGGAGCTCGGGAAGAGCGGCTCCGCCCTCCGCTGGACTCGGCGCTG 265
QY 80 AspGlnGlnLeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGlu 99
Db 266 GCGTTGCGGCTCGGGCTGCAGACGCGCGGCTGGCGGCCCTGTATCCGAGACCTGGAG 325
QY 100 -----PheGlnGluArgLeuGlnGluAspThrPheGlyGlyAsnLeuGly 115
Db 326 AGGGCCCGTTCCTCCGCGGCGGCTGACCGAGCGACACGAGCTTCGTGGGAACCGTGGGA 385
QY 116 Arg---AlaValPheGlnAlaLysLysArgValLeuGluProLeuGlyLeuValGluGlu 134
Db 386 GCGTTGGGAAGCTCGAGCGGCTCCCGCATY----- 418
QY 135 GlyAlaLysThrAlaProGlyLysLysArgProValGluGlnInsProGlnGluProAsp 154
Db 419 GGAATTCGCGAGTCCCGGGAGGAG-----GCCCCAGAT 454
QY 155 SerSerSerGlyLeuGlyThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGly 174
Db 455 ACTCGAGCGCCACGGGAGGAGGACCTCAGGCCACAGTAGTGGGAAGCCGAGATGGCC 514

```

175  GlnThrGlyAspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThr 194
    |||
    |||
    |||
515  TCTCGGGGTGTGGCGCTGCT-----CCAGTGGGGAGGAAAAAAGATG 562
    |||
    |||
    |||
195  ProAlaAlaValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsn 214
    |||
    |||
    |||
563  -----ACAGAGAAGCATGTGGGGCAGGGTCTCCCGCGACA----- 598
    |||
    |||
    |||
215  AsnGluGlyAlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrp 234
    |||
    |||
    |||
599  -----GTGGCGCAGCATGGAT----- 613
    |||
    |||
    |||
235  LeuGlyAspArgValIleThrThrSerThrArgThrTrpAlaLeuProThrThrAsnAsn 254
    |||
    |||
    |||
614  -----ACCTAGACAGCGTCCAGCTAAAGCTAGACACCATGAAT--- 652
    |||
    |||
    |||
255  HisLeuTyrLysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPhe 274
    |||
    |||
    |||
653  -----GCACAGGCTGACAGGGCATATCTC 676
    |||
    |||
    |||
275  GlyTyrSerThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerPro 294
    |||
    |||
    |||
677  AGGCTTTTCAGCAAGTTTGGCAGTTG-----CGACTTCAC---CACTTAGAGCGC 724
    |||
    |||
    |||
295  ArgAsp-----TrpGlnArgLeuIleAsnAsnAsnTrp 305
    |||
    |||
    |||
725  CGGAACCTCCTCATCCAGGACATTCGGGGCTTCGGGGCGAAGCTTTTCAGAACCAT--- 781
    |||
    |||
    |||
306  GlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLysGluValThr 325
    |||
    |||
    |||
782  -----CCCCAGCTGTCAGCTTT-----CTGAATACCAAGATAAGAG- 820
    |||
    |||
    |||
326  ThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGlnValPheSer 345
    |||
    |||
    |||
821  -----GTATTGAGCTATTGAAACAGACTGGAGGTGGGAAGAACTTGGCCTTGCC 868
    |||
    |||
    |||
346  AspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCysLeuProPro 365
    |||
    |||
    |||
869  AGGTGGGCTACAAAATCAAGTTCTACTTTGGCCGA-----AACCCC 910
    |||
    |||
    |||
366  PheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsnAsnGlySer 385
    |||
    |||
    |||
911  TATTTCCTCAAAACAGGTGCTCAATTAAGGAATATGCG-----TGTGGTCCATCCGGT 961
    |||
    |||
    |||
386  GlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGlnMetLeuArg 405
    |||
    |||
    |||
962  CAAGTGGTGTCTCGTTTCAGCT-----CCAAATCCAGTGTCTCCAGGCCATGATCTACAA 1015
    |||
    |||
    |||
406  Thr-----GlyAsnAsnPheThrPheSerTyrThrPheGluGluVal 419
    |||
    |||
    |||
1016  TCCCTAAGCAAGGAAACCCAGAAACCAATGCTAGCTCTTTGGGTGGTTTCAAT--- 1072
    |||
    |||
    |||
420  ProPheHisSerSerTyrAlaHisSerGlnSerLeu-----AspArgLeu 434
    |||
    |||
    |||
1073  -----CACAGCTCTATTGAGTCTGACACAGATTGTTGAGATAATTAATGAGGACCTGTGG 1126
    |||
    |||
    |||
435  MetAsnProLeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGly 454
    |||
    |||
    |||
1127  CCCAATCTCTCTA-----CAGTACTACCTGATCAGTGAAGAAGCCGCTGGAGAGAAAGGA 1180
    |||
    |||
    |||
455  SerAlaGlnAsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGln 474
    |||
    |||
    |||
1181  AAGGAAGAA-----AGGCCAGGTCTCCAGAAAACCTGAGC----- 1213
    |||
    |||
    |||
475  ProLysAsnTrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThr 494
    |||
    |||
    |||
1214  -----CCTGCACCTGCGTGGAGGCGCCCACTGATCCGACACAGATCTC 1258
    |||
    |||
    |||
495  AspAsnAsnAsnSerAsnPheThrTrpThr 504
    |||
    |||
    |||
1259  CCTACCACCTCT-----TGCTGGACC 1279
    |||
    |||
    |||

```

```

RESULT 8
CA356072/c 640 bp mRNA linear EST 05-NOV-2002
LOCUS 628112 NCCWA 1RT Oncorhynchus mykiss cDNA clone 1RT91G17_A_D09 5',
DEFINITION mRNA sequence.
ACCSSION CA356072
VERSION CA356072.1 GI:24601259
KEYWORDS EST.
SOURCE Oncorhynchus mykiss (rainbow trout)
ORGANISM Oncorhynchus mykiss
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
1 (bases 1 to 640)
Rexroad, C.E. and Keeler, J.W.
Sequence analysis of a rainbow trout normalized cDNA library
Unpublished
Contact: Rexroad CE
USDA, ARS, National Center for Cool and Cold Water Aquaculture
11876 Leetown Road, Kearneysville, WV 25430, USA
Tel: 304 724 8340 x2129
Fax: 304 725 0351
Email: crexroad@nccwa.ars.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified by
cross match v0.990329.
Seq primer: AGCGATAACAATTTTCACACAGGA.
FEATURES
Location/Qualifiers
1..640
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/clone="1RT91G17_A_D09"
/tissue_type="pooled"
/lab_host="DH108"
/clone_lib="NCCWA 1RT"
/notes="vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from brain, gill, liver,
spleen, muscle, and kidney."
BASE COUNT 136 a 185 c 198 g 121 t
ORIGIN
Alignment Scores: 0.818 Length: 640
Pred. No.: 125.00 Matches: 63
Score: 37.05% Conservative: 30
Percent Similarity: 25.10% Mismatches: 71
Best Local Similarity: 3.13% Indels: 88
Query Match: 14 Gaps: 13
DB:
US-09-807-802A-13 (1-736) X CA356072 (1-640)
Qy 124 ArgValLeuGluProLeuGlyLeuValGluGlyAlaLysThrAlaProGlyLysLys 143
|||
|||
|||
Db 602 AGACTGTTCCTCCCTTGGTCCCATTTGCTCCGGGCGAGACCGTCCAGGT----- 549
|||
|||
|||
Qy 144 ArgProValGlnSerProGlnGluProAspSerSerGlyIleGly---LysThr 162
|||
|||
|||
Db 548 ATGCCGCTGGAGAGATTCCAGCAGGTCCGGGCTCCAGCAGGTCCAGGCATGCCCTG 489
|||
|||
|||
Qy 163 GlyGln---GlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGluSer 181
|||
|||
|||
Db 488 GGACCTCTGAAGTCCAGCTGCCCTTCCTTC-----TCGCCAGCT 444
|||
|||
|||
Qy 182 ValProAsp-----ProGlnProLeuGlyGluPro----- 191
|||
|||
|||
Db 443 GCACAGGAACACACAGGCTTACAGGACCAACCCGGGAGCCAGGTTGCCAGGTCCGAG 384
|||
|||
|||
Qy 192 -----ProAlaThrProAlaAlaValGly---ProThrThrMet 203
|||
|||
|||
Db 383 TAGCCAGCAGGTCCGGGTGGTCCGCGAGGGGCTCTGTCTCGGCTGTCCACAGCTCT 324
|||
|||
|||
Qy 204 AlaSerGlyGly---GlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGly 222
|||
|||
|||

```

```

Db 333 TCCCTGGGGCCAGGTCAGGCGGGGCGCTCGGGCCGGGCTCACCTGGGGTA--- 267
Qy 223 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeu----- 235
Db 266 -----CCGGGCTCTCCCTGCCACTGACACAGGTCGCTCTTCACAGAGTAGGGCTGGTAC 213
Qy 236 ---GlyAspArgValIleThr-ThrSerThrArgThrTrpAlaLeuProThrTyrAsnAs 254
Db 212 TGGGGGGAGGCGCTTCAATACCTCTCTCA-----CTATATATGAA 174
Qy 254 nHisLeuTyrLysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPh 274
Db 173 CACATACATGAACAGCTCTCA----- 152
Qy 274 eGlyTyrSerThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerPr 294
Db 151 -----AGCCC 147
Qy 294 oArgAspTrpGlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPh 314
Db 146 ACCAGGAGG-AGGAGGATGCTCAATACT-----CTTAC 115
Qy 314 eLysLeuPheAsnIleGlnValLysGluValThrThrAsnAspGlyValThrIleAl 334
Db 114 TTCCATTTTCAGATGCGAGTGAGTGAGTTCAGAGTTCACAGAGGGTAAGAAATCTTCA 55
Qy 334 aAsnAsnLeuThrSerThrValGlnValPhe 344
Db 54 GCACCACTGTCTCTCTACTGCTGTCAGTTTAT 24

```

RESULT 9

CA363789/c 638563 NCCWA 1RT Oncorhynchus mykiss cDNA clone 1RT18G06_C.D03
LOCUS 5', mRNA sequence.

ACCESSION CA363789.1 GI:24672376

VERSION EST.

SOURCE Oncorhynchus mykiss (rainbow trout)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.

1 (bases 1 to 701)

REFERENCE Rexroad, C.E. and Keele, J.W.

AUTHORS Sequence analysis of a rainbow trout normalized cDNA library

TITLE Unpublished

JOURNAL

COMMENT Contact: Rexroad CE

USDA, ARS, National Center for Cool and Cold Water Aquaculture

11876 Leetown Road, Kearneysville, WV 25430, USA

Tel: 304 724 8340 x2129

Fax: 304 725 0351

Email: crexroad@nccwa.ars.usda.gov

Single pass sequencing. Bases called with phred v0.020425.c and

trimmed with the aid of the trim_alt option. Vector identified by

cross_match v0.990329.

Seq primer: ACGGATAACAAATTCACACAGA.

Location/Qualifiers

1..701

/organism="Oncorhynchus mykiss"

/mol_type="mRNA"

/db_xref="taxon:8022"

/clone="1RT18G06_C.D03"

/cissue_type="pooled"

/lab_host="DH10B"

/clone_lib="NCCWA 1RT"

/note="Vector: pCMV SPOT6; Site 1: NotI; Site 2: SalI;

Library made from pooled tissue from brain, gill, liver,

spleen, muscle, and kidney."

150 a 194 c 231 g 126 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 0.927 Length: 701
Score: 125.00 Matches: 63
Percent Similarity: 37.05% Conservative: 30
Best Local Similarity: 25.10% Mismatches: 71
Query Match: 3.13% Indels: 88
DB: 14 Gaps: 13

US-09-807-802A-13 (1-736) x CA363789 (1-701)

```

Qy 124 ArgValLeuGluProLeuGlyLeuValGluGluAlaLysThrAlaProGlyLysLys 143
Db 585 AGACCTGTTTCCCTCCCTTGGTCCATTGCTCCGGCAGACCGTGAGGTCCAGGT----- 532
Qy 144 ArgProValGluGlnSerProGlnGluProAspSerSerSerGlyIleGly-----LysThr 162
Db 531 ATGCCGGTGAAGAGATTCCAGCAGAGTCCGGGGCTTCCAGCAGGTCCAGCATGCCCTGTG 472
Qy 163 GlyGln---GlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGluSer 181
Db 471 GGACCTGAAGTCAGCTGCCCCACCTTTCCTTC-----TCGCCAGCT 427
Qy 182 ValProAsp-----ProGlnProLeuGlyGluPro----- 191
Db 426 GCACCAAGAAACACCAAGCTTACCAAGGACCACCCGGGGAGCCAGGTTTGCACAGTGGCGAG 367
Qy 192 -----ProAlaThrProAlaAlaValGly---ProThrThrMet 203
Db 366 TAGCACACAGTCCGGGTGTCGGGACCGGCTCTGTCGGGCTCTCCACAGTCTTT 307
Qy 204 AlaSerGlyGly---GlyAlaProMetAlaAspAsnGluGlyAlaAspGlyValGly 222
Db 306 TCCCTGGGGGGCAGGTGGCCAGGGGGCCCTCTGGCGGGCTCACACAGTAGTGGCTGTAC 196
Qy 223 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeu----- 235
Db 249 -----CCGGGCTCTCTCTGCACTGACACACCACTGGCTCTTACAGAGTAGTGGCTGTAC 196
Qy 236 ---GlyAspArgValIleThr-ThrSerThrArgThrTipAlaLeuProThrTyrAsnAs 254
Db 195 TGGGGGAGGCGCTTCAATACCTTCTTCA-----CTATATATGAA 157
Qy 254 nHisLeuTyrLysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPh 274
Db 156 CCACATACATGAACAGCTGTCA----- 135
Qy 274 eGlyTyrSerThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerPr 294
Db 134 -----AGCCC 130
Qy 294 oArgAspTrpGlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPh 314
Db 129 ACCAGGAGG-AGGAGGATGCTCAATACT-----CTTAC 98
Qy 314 eLysLeuPheAsnIleGlnValLysGluValThrThrAsnAspGlyValThrIleAl 334
Db 97 TTCCATTTTCAGCATGCAAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGT 38
Qy 334 aAsnAsnLeuThrSerThrValGlnValPhe 344
Db 37 GCACCACTGTCTCTCTACTGCTGCTGCTTAT 7

```

RESULT 10

CA378088/c

LOCUS

DEFINITION

658613 NCCWA 1RT Oncorhynchus mykiss cDNA clone 1RT42114_C.E07 5',

mRNA sequence.

ACCESSION CA378088

VERSION CA378088.1

KEYWORDS

SOURCE

ORGANISM

Oncorhynchus mykiss (rainbow trout)

Oncorhynchus mykiss

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.

1 (bases 1 to 703)
Rexroad, C.E. and Keele, J.W.

Unpublished
Sequence analysis of a rainbow trout normalized cDNA library

CONTACT: Rexroad CE
USDA, ARS, National Center for Cool and Cold Water Aquaculture

11876 Leetown Road, Kearneysville, WV 25430, USA

Tel: 304 724 8340 x2129

Fax: 304 725 0351

Email: crexroad@nccwa.ars.usda.gov

Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified by
cross match v0.990329.

Seq primer: AGCGGATACCAATTTCACACAGGA.

FEATURES

Location/Qualifiers

1..703

/organism="Oncorhynchus mykiss"

/mol_type="mRNA"

/db_xref="taxon:8022"

/clone="1RT42114 C.E07"

/tissue_type="pooled"

/lab_host="DH10B"

/clone_lib="NCCCWA lrt"

/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from brain, gill, liver,
spleen, muscle, and kidney."

146 a 195 c 235 g 127 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 0.93 Length: 703

Score: 125.00 Matches: 54

Percent Similarity: 41.04% Conservative: 17

Best Local Similarity: 3.12% Mismatches: 55

Query Match: 3.13% Indels: 47

DB: 14 Gaps: 11

US-09-807-802A-13 (1-736) x CA378088 (1-703)

QY 124 ArgValLeuGluProLeuGlyLeuValGluGlyAlaLysThrAlaProGlyLysLys 143

DB 578 AGACCTGTTTCCCTCCCTTGGTCCCTTCCCTGCGGAGCCAGTCCAGTCCAGT 525

QY 144 ArgProValGluGlnSerProGlnGluProAspSerSerGlyGlyGlyLysThr 162

DB 524 ATCCCGGTGGAGAGATTCCAGAGTCCGGGCTTCCAGCAGTCCAGGATGCCCTG 465

QY 163 GlyGln---GlnProAlaLysLysArgLeuAsnPhesGlyGlnThrGlyAspSerGluSer 181

DB 464 GGACCTGAAGTCCAGTCCCTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC 420

QY 182 ValProAsp-----ProGlnProLeuGlyGluPro----- 191

DB 419 GCACCAAGAACACCAAGGCTTACCAGGACCAACCCGGGGGAGCCAGTTCAGGAG 360

QY 192 -----ProAlaThrProAlaAlaValGly---ProThrThrMet 203

DB 359 TAGCCAGCAGGTCCGGTGGTCCGGCAGGCGCTCTGTCGGGCTGTCCACAGCTCTT 300

QY 204 AlasSerGlyGly---GlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGly 222

DB 299 TCCCTGGGGGGCCAGGTGGCCAGGCGGCGCTCTGTCGGGCGGCTCACCTGGGTA 243

QY 223 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeu----- 235

DB 242 -----CCGGGCTCTCTCTGCACTGACACCACTGGCTCTTCACAGTAGGCTGGTAC 189

QY 236 ---GlyAspArgValIleThr-ThrSerThrArgThrTrpAlaLeuProThrTyrAsnAs 254

DB 188 TGGGGGAGGCCCTTCACTTCTTCTTCA-----CTATATATGAA 150

QY 254 nHisLeuTyrLysGlnIleSerSerAlaSerThrGly 266

DB 149 CCCTACCATGACAGCTGTCAAGCCCAAGGAGGA 113

RESULT 11

BU139005/c

LOCUS

DEFINITION

Caenorhabditis elegans cDNA clone yk1141g12 3', mRNA sequence.

ACCESSION

BU139005

VERSION

BU139005.1 GI:18299171

KEYWORDS

EST.

SOURCE

Caenorhabditis elegans

ORGANISM

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea

REFERENCE

1 (bases 1 to 745)

AUTHORS

Kohara, Y., Shin-i, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.

TITLE

A complementary view of the C.elegans genome

JOURNAL

Unpublished

COMMENT

Contact: Tadasu Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tahini@genes.nig.ac.jp.

Location/Qualifiers

1..745

/organism="Caenorhabditis elegans"

/mol_type="mRNA"

/strain="N2"

/db_xref="taxon:6239"

/clone="yk1141g12"

/sex="hermaphrodite"

/tissue_type="whole animal"

/dev_stage="L1"

/clone_lib="unpublished oligo-capped cDNA library, C.

elegans L1 stage"

75 a 212 c 238 g 216 t 4 others

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 1.01 Length: 745

Score: 125.00 Matches: 56

Percent Similarity: 35.11% Conservative: 10

Best Local Similarity: 29.79% Mismatches: 72

Query Match: 3.13% Indels: 50

DB: 12 Gaps: 9

US-09-807-802A-13 (1-736) x BU139005 (1-745)

QY 87 AspAsnProTyrLeuArgTyrAsnHisAla---AspAlaGluPhesGlnGluArgLeu--- 104

DB 678 GAAACCCGGATCCCAAGGAGGATCATGCTTNGATGCTTCCTCCAGGAGCCGCTGGAC 619

QY 105 -----GlnGluAspThrSerPheGly 111

DB 618 CAGCCGGAAACCCAGGAAGCCAGGAGCCAGGAGCGTCCAGGAGCAGCTGGACTTCCAA 559

QY 112 GlyAsnLeuGlyArgAlaValPheGlnAlaLysLysArgValLeu---GluProLeuGly 130

DB 558 GGAACCCAGGAGCGCCACAGCAGCCATGTGAGCAATCACCCACACCATGCAAG 499

QY 131 LeuValGluGluGlyAlaLysThrAlaProGlyLysLysArgProValGluGln----- 148

DB 498 CCATGCCCAAGGAGCCAGCTGGAGCCCGGAGGAGCCCGGAGGAGGAGGAGGAGGAG 439

QY 149 SerProGlnGluProAspSerSerGlyIleGlyThrGlyGlnGlnProAlaLys 168

DB 438 GCACCAAGGAGCCCGGACAGGATCAGGAGCCCGGAGGAGGAGGAGGAGGAGGAGGAG 385

QY 169 LysArg-----LeuAsnPhgGlyGlnThrGly----- 177
 Db 384 CCAAGGAGAGCCCGAGGAGCCCGAGGAAACCCAGGACAAAGCCGAGCCCGAGGACCA 325
 QY 178 -----AspSerGluSerValPro----- 183
 Db 324 GGNATCCGAGCCTCAATCCAGAGTCTTCCAGGAGCCCGAGGACCAAGGACCAAGGA 265
 QY 184 AspProGlnProLeuGlyGluPro-----ProAlaThrProAlaVal 198
 Db 264 CCACGAGGACCGAGTCCCGAGGAGCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 205
 QY 199 GlyProThrMetAlaSerGlyGlyAlaProMetAlaAspAsnGluGlyAla 218
 Db 204 GGACCAAGAGGACCATCAGGAGCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 145
 QY 219 AspGlyValGlyAsnAlaSerGly 226
 Db 144 CCAGGACAGCCAGGACAACTGGA 121

RESULT 12

BG207226 754 bp mRNA linear EST 21-APR-2001
 LOCUS RST26694 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
 DEFINITION BG207226

VERSION BG207226.1 GI:13728913

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 754)
 Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,
 Cain, S., Leventhal, C., Thornton, M., Ramachandran, R., Whittington, J.,
 Lerner, L., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith
 E., Veloso, N., Klika, A., Hess, J., Cothren, K., Lo, K., Offenbacher
 J., Danzig, J. and Ducar, M.
 Creation of genome-wide protein expression libraries using random
 activation of gene expression

TITLE

JOURNAL Nat. Biotechnol. 19 (5), 440-445 (2001)

MEDLINE 21227151

PUBMED 11329013

COMMENT

Contact: Scott J. Cain
 Athersys, Inc.
 3201 Carnegie Ave, Cleveland, OH 44115, USA
 Tel: 216 431 9900
 Fax: 216 361 9596

Email: scain@athersys.com

High quality sequence stop: 326.

Location/Qualifiers

source

1. 754

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/cell_line="HT1080"

/clone_lib="Athersys RAGE Library"

/notes="See 'Creation of Genome-wide Protein Expression
 Libraries using Random Activation of Gene Expression',
 Nature Biotechnology, in press. Note that even though the
 cell type indicated is HT1080, since a random activation
 method was used, these sequence tags are not necessarily
 expressed in HT1080 under normal circumstances."

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 1.02 Length: 754

Score: 125.00 Matches: 50

Percent Similarity: 33.06% Conservatve: 31

Best Local Similarity: 20.41% Mismatches: 68

Query Match: 3.13% Indels: 96

DB: 10 Gaps: 11

US-09-807-802A-13 (1-736) x BG207226 (1-754)
 QY 444 TyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuPhe 463
 Db 206 TTTTATCTCAACCAACGAA-----CGAGGAGGTAAATAATCTAAGGAGCACTGTC 256
 QY 464 SerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyProCys 483
 Db 257 CAAAGAGTACAGAAACAGGCTCGCCGTGGAATAGGAACTGGATG----- 304
 QY 484 TyrArgGlnGlnArgValSerLysThrAspAsnAsnAsnSerAspPheThrTrp 503
 Db 305 -----ACTCGACAGGCAAGCGAGAGTCTACAGATGTTGATGCGCTTGGCCCTGCTCAGTGTCA 358
 QY 504 ThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThrAla 523
 Db 359 CAAGGA-----AATCTGATTTTCCCTGGGGTTTCACTG 388
 QY 524 MetAlaSerHisLysAspGluAspLysPhePheProMetSerGlyValMetIlePhe 543
 Db 389 TTGGCTCTTGATGACGAGCCAGGATTCCTGGATGGCTTGGCCCTGCTCAGTGTG 448
 QY 544 GlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAspGlu 563
 Db 449 CGACGCCAGACTCTG-----GACGCATACCTATCCGGGCC-----CGCCCACTG 550
 QY 623 Pro-----HisThrAspGlyHisPheHis----- 630
 Db 551 CCGAAAAACCAAGCTTACAGACACTGCACCAACCCCTGTCAAGTAGGCTTACTACAAA 610
 QY 631 -----ProSerProLeuMetGlyGlyPhe 638
 Db 611 CGGAAGCCGCGCGAGGCAAAAAACCAACCCGATCGCCGACCCCTTAGCGCA----- 664
 QY 639 GlyLeuLysAsnProProGlnIleLeuIleLysAsnThr-----ProVal 654
 Db 665 TAAATATCCGACCCAGCCAGATCAAGACAAACCCGACCGGACCGAGCAACCCACG 724
 QY 655 ProAlaAsnProPro 659
 Db 725 CCTGGCAACCCCTCCG 739

RESULT 13

LOCUS

DEFINITION

AK014322

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Scott J. Cain

Athersys, Inc.

3201 Carnegie Ave, Cleveland, OH 44115, USA

Tel: 216 431 9900

Fax: 216 361 9596

Email: scain@athersys.com

High quality sequence stop: 326.

Location/Qualifiers

source

1. 754

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/cell_line="HT1080"

/clone_lib="Athersys RAGE Library"

/notes="See 'Creation of Genome-wide Protein Expression
 Libraries using Random Activation of Gene Expression',
 Nature Biotechnology, in press. Note that even though the
 cell type indicated is HT1080, since a random activation
 method was used, these sequence tags are not necessarily
 expressed in HT1080 under normal circumstances."

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 1.02 Length: 754

Score: 125.00 Matches: 50

Percent Similarity: 33.06% Conservatve: 31

Best Local Similarity: 20.41% Mismatches: 68

Query Match: 3.13% Indels: 96

DB: 10 Gaps: 11

AK014322 2019 bp mRNA linear HTC 05-DEC-2002
 Mus musculus 14, 17 days embryo head cDNA, RIKEN full-length
 enriched library, clone:322402004 product:similar to LAP-4, PROTEIN
 (LYMPHOID NUCLEAR PROTEIN RELATED TO AP4) (Homo sapiens), full
 insert sequence.

AK014322

VERSION AK014322.1 GI:12852094

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1

AUTHORS Carninci, P. and Hayashizaki, Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

147 GluGlnSerProGlnGluPuroAaspSerSerSerglyleGlyLysThr----- 162
:: ||| ||||| ::
459 GATGCGGAGCCAGGGAGTTCAGACAATTCGAGCATTCGAACACTTCAACTTCGGAAGACGAC 518

163 -----GlyGlnGlnProAlaLysLysArgLeuAsnPhe 173
:: ||| ||||| ::
519 CTTAAAGCTAAGCAGTGATGACGAAGAGGTGAGCAGCAGCAGCCCCAGAGAACAGACTCTC 578

174 GlyGlnThrGlyAaspSerGluSerValProAaspProGlnProLeuGlyGluProProAla 193
||| |||||
579 CGCGCTTAGCTGACAGCTCCGTGGTTTCAGCACACCACCACTGTAGAGTTCGGCGCCATCC 638

194 ThrProAlaValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAsp 213
:: ||||| ::
639 AGCAAGGCC-----GGCGCAGCAGCAGCAGCAGCGCGCGCAGAGCTCTCTCCAGCGAC 692

214 -----AsnAsnGluGlyAlaAaspGlyValGlyAsnAlaSerGlyAsnTrp----- 228
:: ||||| ::
693 TCAGAGAGCACTTCGCGCTCCGACTCCGACGAGCAGGAGCAGCTCCAGCTCCAGCGAGAGC 752

229 -----HisCysAaspSerThrTrpLeuGlyAaspArgValIleThrThr 242
||| |||||
753 GAGGGCAGCAAGCCTCCGCACCTGCTCCAGC-----CCCGAGGCCGCAACCTGGC 800

243 SerThrArgThrTrpAlaLeuProThrTyzAsnAsnHisLeuTyzLysGln----- 259
||| ||||| ::
801 TCCTCTAACAAAGTGGCAGCTGGATAATGGCTTAACAAAGTTAATCTCTCAAGCCGCCCA 860

260 -----IleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyzPheGlyTyzSer 277
||| ||||| ::
861 ATTCTGATCAAATGAAGCCACCGCGCCCGGAAGGAATCAATACTACACGCCGCCCGGTG 920

278 ThrProTrpGlyTyzPheAaspPheAsnArgPheHis-----CysHisPheSerProArg 295
||| ||||| ::
921 AAGATCAAGGGCNA---GACTCGGGAAACTTCCTGAGATCTGCCNAGCCACCTGAGG 977

296 AspTrpGlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLys 315
:: ||| ||||| ::
978 GACAAGGAGCTCAAGACCACTCTGCAAGGAGGAGCAGAGCGCCGAGGACCGCAACAAGGCC 1037

316 LeuPheAsnIleGlnValLysGluValThrThraAsnAspGlyValThrThrIleAlaAsn 335
||| ||||| ::
1038 CCTGGCAGCAAAAGTGTAAAGCAGAGTCCCACCGCTGCTGGCTGTGACCGCAGCT 1097

336 AsnLeuThrSerThrValGlnValPheSerAaspSerGluTyzGlnLeuProTyzValLeu 355
||| |||||
1098 GCCCTCCCGCCCGGGTGCCAGCGCACCCACGAGAGCGCGCTGCGCCC---ACCGCG 1154

356 GlySerAlaHisGlnGlyCysLeuProProPheProPheProAlaAspValPhe----- 371
||| ||||| ::
1155 AGGTCAGCAGG-CAAGAAGCCCAACCCCGCGCACCGAGAGGACTCTACCGCGGAGCGGTGC 1213

372 ---MettileProGlnTyzGlyTyzLeuThrLeuAsnAsnGlySerGlnAlaValGlyArg 390
||| ||||| ::
1214 CAACTGCCACCGGCTGAGGA-----GCCAGTGGCCCGACAGACAGCTGGGGCG 1361

391 SerSerPheTyzCysLeuGluTyzPheProSerGlnMetLeuArgThrGlyAsnAsnPhe 410
||| ||||| ::
1262 AGCGTGGTGGGACCCCTTAGAG-----CCCCCAAAAACAGACCTCCGCGTAAACAAC--- 1312

411 ThrPheSerTyzThrPheGluGluValProPheHisSerSerTyzAlaHisSerGlnSer 430
||| |||||
1313 -----AGGACAGGCCCAACCGCAGGAGCTCGCTCTCTCCGTGACCTGTGAGAGCGG 1363

431 LeuAspArgLeuMetAsnProLeuIleAaspGlnTyzLeuTyzLeuAsnArgThrGln 450
||| ||||| ::
1364 CGCACGAGGGGGCTGAGCAGAACTCGTCCCAAAATCCAAGGAGTTATTATGAACAGAGATCT 1423

451 AsnGlnSerGlySerAlaGlnAsnLysAspLeuPheSerArgGlySerProAlaGly 470
||| |||||
1424 TCGTCTTCCTCTCTCTCTCGGACTCAGACCTG----- 1456

471 MetSerValGlnProLysAsnTrpLeu-----ProGlyProCysTyzArgGlnGln 487

| | | | |
|--|--|--|-----------------|
| Db | 1457 | GAGTCTAGACGAGAGGAGTACGTCTGTCTCAAGGCTCCGACCACCACTGGGACCGAGCAG | 1516 |
| Qy | 488 | ATgvalSerLysThrLysThrAspAsnAsnAsnSerAsnPheThrTrpThrGlyAlaSer | 507 |
| Db | 1517 | CGGCTGAAGGCGCCGACGAGCAACAACAGCAAC | 1555 |
| Qy | 508 | LysTyrAsnLeuAsnGlyArgGlusErIle | 525 |
| Db | 1556 | AGCAACAGCAGCAGCACCTCCCGGGCTCTGTGGGCTCATCAACGCCAGCACCACTGAC | 1615 |
| Qy | 526 | SezHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIlePheGlyLys | 545 |
| Db | 1616 | ATTGCCAAGGAGCTAGAGGACAGTCTTACACGCTG | 1666 |
| Qy | 546 | GlusErAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAspGlu | 563 |
| Db | 1667 | -----AATGAACCTACTATCTCCGCTAAAGGACAGCATGAG | 1702 |
| RESULT | 14 | | |
| LOCUS | BQ440075 | 925 bp mRNA linear | EST 24-MAY-2002 |
| DEFINITION | AGENECOURT_7831393 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6152062 | | |
| ACCESSION | BQ440075 | | |
| VERSION | BQ440075.1 | GI:21179151 | |
| KEYWORDS | EST. | | |
| SOURCE | Homo sapiens (human) | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | |
| AUTHORS | 1 (bases 1 to 925) | | |
| TITLE | NIH-MGC http://imgc.nci.nih.gov/. | | |
| JOURNAL | National Institutes of Health, Mammalian Gene Collection (MGC) | | |
| COMMENT | Unpublished Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation by: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM13489 row: e column: 23 High quality sequence stop: 480. Location/Qualifiers | | |
| FEATURES | | | |
| SOURCE | 1..925 | /organism="Homo sapiens" | |
| | | /mol_type="mRNA" | |
| | | /db_xref="taxon:9606" | |
| | | /clone="IMAGE:6152062" | |
| | | /issue_type="retinoblastoma" | |
| | | /lab_host="DH10B (phage-resistant)" | |
| | | /clone_lib="NIH MGC 67" | |
| | | /note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.75 kb. Library constructed by Life Technologies." | |
| BASE COUNT | 190 a 289 c 309 g 136 t | 1 others | |
| ORIGIN | | | |
| Alignment Scores: | | | |
| Pred. No.: | 1 66 | Length: | 925 |
| Score: | 124.00 | Matches: | 59 |
| Percent Similarity: | 30.80% | Conservative: | 22 |
| Best Local Similarity: | 22.43% | Mismatches: | 100 |
| Query Match: | 3.11% | Indels: | 82 |
| DB: | 13 | Gaps: | 12 |
| US-09-807-802A-13 (1-736) x BQ440075 (1-925) | | | |
| Qy | 63 | GluproValAsnAlaAlaAspAlaAlaLeuGluHisAspLysAlaTyrAspGlnGln | 82 |

```
Db 180 GAGCCAGCTGAGCTAGCTGGGGCTACCCAGCTTGGGAAGAGAT- |||||
Qy 83 LeuLysAlaGlyAspAsnProTyrLeuArgTyrAsnHisAlaAspAlaGluPheGlnGlu 102
Db 231 CTTTCCAGGCCCCGAG- |||||
Qy 103 ArgLeuGlnGluAspThrSerPheGlyGlyAsnLeuGlyArgAlaValPheGlnAlaLys 122
Db 272 -----
Qy 123 LysArgValLeuGluProLeuGlyLeuValGluGluGlyAlaLysThrAlaProGlyLys 142
Db 273 -----CTCTCAGAGCCGGACTCCGAGAGAGGAAGAAGATGAG 308
Qy 143 LysArgProValGluGlnSerProGlnGluProAspSerSerSerGlyIleGlyLysThr 162
Db 309 GAGGAGGAGAGAGAGACCACTGACGATCTCTGATATGATCTCTGCG- 356
Qy 163 GlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGluSerVal 182
Db 357 -----TACAAGGTGAAGCAGCGCTTGGCGGGCGCGTGGTGGCCCATCCCGCGGCCC 410
Qy 183 Pro-----AspProGlnProLeuGly 189
Db 411 CCCCGTGCAGCCAGCCCGCCCGCCAGCTTGCAGCTTGTGGCGCTCACCCCTTGGG 470
Qy 190 Glu-----ProProAlaThrProAlaAla-----ValGlyProThrThrMetAlaSerGly 206
Db 471 GAGGCCCCACCGGGAACCCACCTGCGCGCTCTGCTGCTGTGTACAGCCCCCCAGGA 530
Qy 207 GlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGlyAsnAlaSerGly 226
Db 531 AGCACCAGCCCTGAAGGAGGCGGCCCTCGGGA-----GGGAGAGGAGGAACCACTCG 584
Qy 227 AsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrSerThrArgThr 246
Db 585 GCTCGGGAGGCGCCAGCACCTTGGCGGGA-----AGAGGAAGAGGAACAGAGGAGGAGG 641
Qy 247 TrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSerAlaSerThrGly 266
Db 642 GAACCTACCCCTGTACCGGAT-----GTGGAGGATTCCTTTCACCAAC 689
Qy 267 AlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTyrPheAspPheAsn 286
Db 690 CTGGGGGGAGGCTGCACCGG-----GGC 713
Qy 287 ArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsnAsnAsnTrpGly 306
Db 714 ACTTTTCATTGC-----TTGCAATGCCCCCGGGGGTGAAGGGGGTTATGGG 761
Qy 307 PheArgPro 309
Db 762 AAAAAACCC 770
```

```
RESULT 15
BI911451
LOCUS 603063370F1 NIH_MGC_118 Homo sapiens cdna clone IMAGE:5212577 5',
DEFINITION mRNA sequence.
ACCESSION BI911451
VERSION BI911451.1 GI:16175192
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 645)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
```

Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LHAM1533 row: 1 column: 18
High quality sequence start: 27
High quality sequence stop: 645.

FEATURES
source

```
1. .645
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5212577"
/tissue_type="leukocyte"
/lab_host="DH10B"
/clone_lib="NIH_MGC_118"
/note="vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV
(destroyed); RNA source leukocytes from anonymous pool of
non-activated adult donors. Library is oligo-dT primed
and directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
1.2-3.3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 027. Note:
this is a NIH MGC Library."
```

BASE COUNT 113 a 226 c 211 g 95 t
ORIGIN

Alignment Scores:
Pred. No.: 1.12 Length: 645
Score: 123.50 Matches: 39
Percent Similarity: 43.97% Conservative: 12
Best Local Similarity: 33.62% Mismatches: 46
Query Match: 3.10% Indels: 19
DB: 4

US-09-807-802A-13 (1-736) x BI911451 (1-645)

```
Qy 114 LeuGlyArgAlaValPheGlnAlaLysArgValLeuGluProLeuGlyLeuValGlu 133
Db 22 TTAGGATCGGCAAGGGTGCAGCGCTGGGGAGCGGACGCTGTCATCCCTTTCTGTGTCAGGA 81
Qy 134 GluglyAlaLysThrAlaProGlyLysLysArg-----ProValGlu 147
Db 82 ACCTCTCATCAGACCGCTGAGGGAAGCGCGCCCGGAGACCCCGCCCGGTCAC 141
Qy 148 GlnSerProGlnGluProAspSer-----SerSerGlyIleGlyLysThr 162
Db 142 ATTCCTCCAGGAAGCGGACTCTATGGGCGGACCCCTTATGGGCGGAGCCTGAGCCGAGGCC 201
Qy 163 GlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGlyAspSerGluSerVal 182
Db 202 GGAGCCAGCCCGCAACCCCTGAACCTCCAGCCAGGGCGCCCGGAGCGAGC----- 252
Qy 183 ProAspProGlnProLeuGlyGluProProAlaThrProAlaAlaValGlyProThrThr 202
Db 253 -----CAGCCCGTGGGCGAGCCCGCCCGCCCGGAGCGGCGCATGAGCGAGCGGTC 303
Qy 203 MetAlaSerGlyGlyAlaProMetAla-----AspAsnAsnGlu 216
Db 304 ATCTGTTCCAGCCGGCCACTGTGATGCTTTATGATGATGCAACAAG 351
```

Search completed: February 15, 2004, 01:27:56
Job time : 4132.43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 14, 2004, 15:19:20 ; Search time 6088.71 Seconds
(without alignments)
4024.639 Million cell updates/sec

Title: US-09-807-802A-15
Perfect score: 3251
Sequence: 1 TAPGKRPVEQSPQEPDSS.....NNGLYTEPRPIGTRYLTRPL 599

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/csm2_1/USPTO_epool/US09807802/runat_11022004_175607_15929/app_query.fasta_1.2389
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=spt -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HSPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09807802 @CEN 1 1 1216 @runat_11022004_175607_15929 -NCPU=6 -ICPU=3
-NO.MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXI=7

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pin.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|--------|---------------|--------|----|------------------------|
| 1 | 3251 | 100.0 | 4718 | 14 | AF063497 Adeno-888 |
| 2 | 3229 | 99.3 | 4683 | 14 | AF028704 Adeno-888 |
| 3 | 2832 | 87.1 | 4722 | 14 | AF028705 Adeno-888 |
| 4 | 2815 | 86.6 | 4726 | 14 | U48704 Adeno-888 |
| 5 | 2801 | 86.2 | 8179 | 6 | AR264580 Sequence |
| 6 | 2764.5 | 85.0 | 4072 | 6 | AR222046 Sequence |
| 7 | 2764.5 | 85.0 | 4072 | 6 | AR205074 Sequence |
| 8 | 2764.5 | 85.0 | 4679 | 6 | AX282480 Sequence |
| 9 | 2764.5 | 85.0 | 4679 | 14 | AF043303 Adeno-888 |
| 10 | 2764.5 | 85.0 | 7327 | 12 | AF369963 Cloning v |
| 11 | 2764.5 | 85.0 | 7557 | 6 | AR222045 Sequence |
| 12 | 2764.5 | 85.0 | 7557 | 6 | AX205073 Sequence |
| 13 | 2764.5 | 85.0 | 8698 | 6 | AR222044 Sequence |
| 14 | 2764.5 | 85.0 | 8698 | 6 | AX205072 Sequence |
| 15 | 2761.5 | 84.9 | 2208 | 14 | AY243021 Non-human |
| 16 | 2735 | 84.1 | 2211 | 14 | AY242997 Non-human |
| 17 | 2722.5 | 83.7 | 4680 | 6 | AY2828767 Sequence 1 |
| 18 | 2722.5 | 83.7 | 4680 | 6 | I62303 Sequence 1 |
| 19 | 2715 | 83.5 | 2217 | 14 | AY243009 Non-human |
| 20 | 2710 | 83.4 | 2217 | 14 | AY243015 Non-human |
| 21 | 2709.5 | 83.3 | 4721 | 14 | AF513851 Adeno-888 |
| 22 | 2706.5 | 83.3 | 4675 | 6 | AX135805 Sequence |
| 23 | 2706.5 | 83.3 | 4675 | 6 | AX286292 Sequence |
| 24 | 2706.5 | 83.3 | 4675 | 6 | BD094552 Method of |
| 25 | 2706.5 | 83.3 | 4675 | 14 | AA2CG J01901 Adeno-888 |
| 26 | 2703 | 83.1 | 2187 | 14 | AY243013 Non-human |
| 27 | 2701.5 | 83.1 | 2196 | 14 | AY243010 Non-human |
| 28 | 2700 | 83.1 | 2217 | 14 | AY243007 Non-human |
| 29 | 2699 | 83.0 | 2217 | 14 | AY243023 Non-human |
| 30 | 2696.5 | 82.9 | 2208 | 14 | AY243008 Non-human |
| 31 | 2696.5 | 82.9 | 2214 | 14 | AY243020 Non-human |
| 32 | 2696 | 82.9 | 2187 | 14 | AY243006 Non-human |
| 33 | 2695 | 82.9 | 2187 | 14 | AY243017 Non-human |
| 34 | 2692 | 82.8 | 2217 | 14 | AY243022 Non-human |
| 35 | 2691.5 | 82.8 | 2196 | 14 | AY243011 Non-human |
| 36 | 2689 | 82.7 | 2187 | 14 | AY243016 Non-human |
| 37 | 2686 | 82.6 | 2187 | 14 | AY243019 Non-human |
| 38 | 2684.5 | 82.6 | 2190 | 14 | AY242998 Non-human |
| 39 | 2684.5 | 82.6 | 2202 | 14 | AY243014 Non-human |
| 40 | 2683.5 | 82.5 | 2202 | 14 | AY243012 Non-human |
| 41 | 2682.5 | 82.5 | 2202 | 14 | AY243004 Non-human |
| 42 | 2679.5 | 82.4 | 2190 | 14 | AY242999 Non-human |
| 43 | 2679 | 82.4 | 4393 | 14 | AF513852 Adeno-888 |
| 44 | 2676.5 | 82.3 | 2193 | 14 | AY243000 Non-human |
| 45 | 2674 | 82.3 | 2187 | 14 | AY243018 Non-human |

ALIGNMENTS

```

AF063497      4718 bp      DNA      linear      VRL 27-APR-1999
LOCUS      Adeno-associated virus 1, complete genome.
DEFINITION      Adeno-associated virus 1, complete genome.
ACCESSION      AF063497
VERSION      AF063497.1 GI:4689096
KEYWORDS      .
SOURCE      Adeno-associated virus 1
ORGANISM      Adeno-associated virus 1
VIRUSES; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
REFERENCE      1 (bases 1 to 4718)
AUTHORS      Xiao, W., Chirmule, N., Berta, S.C., McCullough, B., Gao, G. and
            Wilson, J.M.
TITLE      Gene therapy vectors based on adeno-associated virus type 1
JOURNAL      J. Virol. 73 (5), 3994-4003 (1999)
MEDLINE      99214338
PUBMED      10196295
REFERENCE      2 (bases 1 to 4718)
AUTHORS      Xiao, W. and Wilson, J.M.
TITLE      Direct Submission
JOURNAL      Submitted (05-MAY-1998) IHGT, University of Pennsylvania, 3601
            Spruce Street, Philadelphia, PA 19104, USA
FEATURES             Location/Qualifiers
     1..4718
     /organism="Adeno-associated virus 1"
     /mol_type="genomic DNA"
     /db_xref="taxon:85106"
     /note="AAV1"
     335..2206
     /codon_start=1
     /product="nonstructural protein"
     /protein_id="AAD27757.1"
     /db_xref="GI:4689097"
     /translation="MPGFYEIVKPSDLDELPGISDFVSVVAKEWELPPDSMD
            LNLISQAPLTAEKQDPLQVWRVSKAPEALFFVQFEKGESYFHLHLVETTVK
            MYLGRFSQIRLQVTVYRIGIEPTLPNNFVTRNGAGGKGVNDECYIPNLLPK
            TOPELQWANTNMEVYSIACLAERKLVAOHLVTSQTEQKENVLPNSDAPVRS
            KTSARVELVMDVRGISTEKQWIOEQAISFNAAASRSQIKALDINAGKIMAL
            TKSAPDYLGPAPPADIKTNRLVRIELNGYBPVAVGSVFLGWAOKRFGKRTIWLFG
            PATTKNTINIAEAIAHVPYGCWNTNENFPNDQVDMKVIWESGKMTAKVESAKA
            ILGGSKRVQDKCSAQIDPTPIVTSNTNMCVIDGNTSTTFEQQQLQDMKFFEL
            TRLEHDFOKTVQKEVFEFFWAQDHTVEAFYVRKGANKRPAPDADKSEPKRA
            CFSVADPSTSDAEGAPVDFADRYQNKSRHAGMLQWLPCKTCERMNQNFICFTHGT
            RDCSECPGVSVSQPVVRKRTYKLCALHLLGRAPIACSAACDLVNVLDLDDCVSEQ"
     2233..4433
     /codon_start=1
     /product="capsid protein"
     /protein_id="AAD27757.1"
     /db_xref="GI:4689097"
     /translation="MAADGYLPDLELDNLSEIREWDLKPGAPKPKANQKQDDGRG
            LVLPGYKYLPGNLDKGPVNAADAALAEHDKAYDQOLKAGDNPYLRYNDAEFQE
            RLQEDTSFGNLRGAVFOAKGVLEPLGLVSGAKTAPGKKEPVQSPOEPDSSGIG
            KTGQPAKRLNFGQGTGDSSEVPDQPLGEPFATPAAGVPTTMASGGGAPADNRGA
            DVGASGNHCHDSTWGLDRVITSTRTWALPTNNHLYKQISSASTGASNDHVFYGY
            STPWGDFDNRPHCFSPRDQWRLLNNNGFRKLNFKLNIQVKEVTINDGVTTIA
            NNLTSVQVFSDESLQPLVLSAHOGLPPFPADVPMIPOGYLTLANGSOAVGRSS
            FLYCLEVPSQMLTGNFTFSYTFEVPFHSSVAHSQSLDRMLNPLIDQYLYLNRTQ
            NQSGAQNKDLLEPSGSPAGMSVQPKMLPGPCVROQVSKTKTDNNNSNFTWTGASK
            YNLNGRESINFGTAWASHKDDKFFPMSGVMIKESAGASNTALDNVMTDEEII
            KATNPVATREFNGTAVNFOSSDPATGDVHAMGALPGVWQDRDVIYLGQPIWAKIPH
            TDGHPSPKLSMFGGLNPPQILINKTEVPANPAPFASATKPAFPIQYSTGVQSVSE
            ISWELQKENSKEWNPVQVTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLRPL"
BASE COUNT      1121 a 1393 c 1273 g 931 t
ORIGIN
Alignment Scores:
Pred. No.:      4,83e-231      Length:      4718
Score:          3251.00      Matches:      599
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      100.00%      Indels:      0
DB:              14      Gaps:      0
US-09-807-802A-15 (1-599) x AF063497 (1-4718)

```

```

1 ThrAlaProGlyLysLysArgProValGluInSerProGlnGluProAaspSerSerSer 20
2634 ACGGCTCTCTGAAAGAAACGTCGGTAGAGCAGTCGCCACAAGAGCAGACTCTCTCTCG 2693
21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAenPheGlyGlnThrGly 40
2694 GGCATCGGCAAGACAGAGCCAGCCGCTTAAAGAAGAGACTCAATTTTGGTCAGACTGGC 2753
41 AspSerGluSerValProAaspProGlnProLeuGlyGlnProAlaLysPheProAlaAla 60
2754 GACTCAGAGTCAGTCCCGCATCCACAACCTCTCGGAGAACCTCCAGCAACCCCGCTGCT 2813
61 ValGlyProThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAaspAenAenGlyGly 80
2814 GTGGGACCTACTACATGGCTTACGGCGGTGGCCACCAATGGCAGACACATAACGAGGC 2873
81 AlaAaspGlyValGlyAenAlaSerGlyAenTrpHisCysAaspSerThrTrpLeuGlyAasp 100
2874 GCCGACGGAGTGGGTAAATGCCTCAGGAATTTGGCATTCCGATTCCACATGGCTGGCGAC 2933
101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTyrAenAenHisLeuTyr 120
2934 AGAGTCATCACACAGCAGCCGACCTGGGCGCTTGGCCACCTACCAATAACCACTCTTAC 2993
121 LysGlnIleSerSerAlaSerThrGlyAlaSerAenAaspAenHisTyrPheGlyTyrSer 140
2994 AAGCAATCTCCAGTGCTTCAACGGGGCCAGCAACGACCACTACTTCTGGCTACAGC 3053
141 ThrProTrpGlyTyrPheAaspPheAenArgPheHisCysHisPheSerProAaspTrp 160
3054 ACCCCCTGGGGGTATTTGATTTTCAACAGATTCCACTGCCACTTTTACCACGCTGACTGG 3113
161 GlnArgLeuIleAenAenAenTrpGlyPheArgProLysArgLeuAenPheLysLeuPhe 180
3114 CAGCGACTCATCAACAACAAATTTGGGGATTCGGGCCCAAGAGACTCAACTTCAAACTCTTC 3173
181 AenIleGlnValLysGluValThrThrAenAaspGlyValThrThrIleAlaAenAenLeu 200
3174 AATCTCCAGTCAAGAGGTCCAGCGCAATGATGGCGTCACACCATCGCTAATAACCTT 3233
201 ThrSerThrValGlnValPheSerAaspSerGluTyrGlnLeuProTyrValLeuGlySer 220
3234 ACCAGACGGTTCAGTCTTCTCGGACTCGGAGTACCAAGCTTCGGTACTGCTCTCGGCTCT 3293
221 AlaHisGlnGlyCysLeuProProPheProAlaAaspValPheMetIleProGlnTyrGly 240
3294 CGGCACCAAGGCTGCTCTCTCCGTCGGCGAGCGTTCATGATTTCCGCAATACCGGC 3353
241 TyrLeuThrLeuAenAenGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
3354 TACCTGAGCTCAACATGGCAGCCAGCCGTCGGAGCTTCATCTCTTTTACTGCTGGAA 3413
261 TyrPheProSerGlnMetLeuArgThrGlyAenAenPheThrPheSerTyrThrPheGlu 280
3414 TATTTCTCTTCTCAGATGCTCAGAACGGGCAACAACATTTTACCTTTCAGCTACACCTTTGAG 3473
281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAaspAtqLeuMetAenPro 300
3474 GAAGTGCCTTTCCACAGCAGCTACCGGCACAGCAGCAGCCGTCGGACCGCTCATGATCT 3533
301 LeuIleAaspGlnTyrLeuTyrTyrLeuAenArgThrGlnAenGlnSerGlySerAlaGln 320
3534 CTCATCGACCAATACCTGTATTACTGAACAGACTCAAAATCATGTCGGGAAGTGCCTCAA 3593
321 AenLysAaspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAen 340
3594 AACAGGACTTGTCTGTTTAGCGTGGTCTCCAGCTGGCATGTCTGTTCAGCCCAAAAAAC 3653
341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAaspAenAen 360
3654 TGGCTACTGTGACCTGTTTATCTGGCAGCAGCGCGCTTTTAAACAAAAACAGACACAAAC 3713

```

```
Qy 361 AenSerAenPheThrTrpThrGlyAlaSerLysTyrAenLeuAenGlyArgGluSerIle 380
Db 3714 AACAGCAATTTTACCTGGACTGGTCTTCAAAATATAACCTCAATGGCGTGAATCCATC 3773
Qy 381 IleAenProGlyThrAlaMetAlaSerHisLysAspGluAenLysPhePheProMet 400
Db 3774 ATCAACCTGGCAGTGTATGGCTTCACAAAGACGACGAGCAAGATGTTCTTCCCATG 3833
Qy 401 SerGlyValMetIlePheGlyGluSerAlaGlyAlaSerAenThrAlaLeuAenAen 420
Db 3834 AGCGGTGTCAATGTTTGGAAAGAGAGAGCGCGGAGCTTCAACACATGCTTGGACAAT 3893
Qy 421 ValMetIleThrAspGluGluGluIleLysAlaThrAenProValAlaThrGluArgPhe 440
Db 3894 GTCATGATTACAGACGAAAGAGAAATTAAGCCACTTAACCTGTGGCCACGAAAGATT 3953
Qy 441 GlyThrValAlaValAenPheGlnSerSerThrAenProAlaThrGlyAenValHis 460
Db 3954 GGGACCGTGGCAGTCAATTTCCAGAGCAGCAGCAGCCTGCGACCGAGATGTGCAT 4013
Qy 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAenArgAenValTrpLeuGlnGlyPro 480
Db 4014 GCTATGGCAGCATTTACCTGGCATGCTGTGGCAAGATAGAGAGTGTACCTGCAGGGTCCC 4073
Qy 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
Db 4074 ATTTGGGCAAAATTCCTCACACAGATGACATTTTCAACCGTCTCTCTTATGGCGGC 4133
Qy 501 PheGlyLeuLysAenProProGlnIleLeuLysAenThrProValProAlaAen 520
Db 4134 TTTGACCTCAGAACCGCTCTCAGATCTCATCAAAACACGCTGTTCTTGGCAT 4193
Qy 521 ProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
Db 4194 CCTCGCGGAGTTCAGTACAAAGTTTGCTTCATTCATCACCAATATCTCCACAGGA 4253
Qy 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAenSerLysArgTrpAenPro 560
Db 4254 CAAGTGAGTGTGAAATTAATGATGGAGCTGCGAAGAAAGAAACAGCAAGCGCTGGAAATCCC 4313
Qy 561 GluValGlnTyrThrSerAenTyrAlaLysSerAlaAenValAenPheThrValAenAen 580
Db 4314 GNAATGCGATACATCTCAATATGCAAAATCTGCAACGTTGATTTTACTGTGGACAAC 4373
Qy 581 AenGlyLeuTyrThrGluProArgProIleGlyThrArgTyrGlyLeuThrArgProLeu 599
Db 4374 AATGACATTATACTGAGCTCGCCCATTTGGCACCCTGATCTTACCGTCCCTG 4430

RESULT 2
AF028704 LOCUS AF028704 4683 bp DNA linear VRL 12-JAN-1998
DEFINITION Adeno-associated virus 6, complete genome.
ACCESSION AF028704
VERSION AF028704.1 GI:2766605
KEYWORDS Adeno-associated virus 6
SOURCE Adeno-associated virus 6
ORGANISM Adeno-associated virus 6
REFERENCE 1 (bases 1 to 4683)
AUTHORS Rutledge, E.A., Halbert, C.L. and Russell, D.W.
TITLE Infectious clones and vectors derived from adeno-associated virus
(JAV) serotypes other than AAV type 2
J. Virol. 72 (1), 309-319 (1998)
JOURNAL JOURNAL
MEDLINE 98080418
PUBMED 9420229
REFERENCE 2 (bases 1 to 4683)
AUTHORS Rutledge, E.A. and Russell, D.W.
TITLE Direct Submission
JOURNAL Submitted (24-SEP-1997) Hematology, Univ. of Washington, Box
357720, Seattle, WA 98195, USA
FEATURES
Location/Qualifiers
1..4683
/organism="Adeno-associated virus 6"
```

```
/mol_type="genomic DNA"
/db_xref="taxon:68558"
/notes="AAV6"
320..2191
/codon_start=1
/evidence=not experimental
/product="nonstructural protein Rep78"
/protein_id="AAB95449.1"
/db_xref="GI:2766606"
/translations="MFGFVEIVIKVPSDLDLHFGISDSFVNVVAEKWBELPPDSMD
LNLIEQAPLTAERKLDQFLVQWRVSKAPEALFFVQFEKESYFHLHIVETVTGVS
MNLGRLEISQIRKLVQTIYRGIERPLNMPAVTKTRNGAGGKNKVVDECVIYNLLPK
TQBELQWATNMEEYISACLNLAERKELVAHDLTHVSQTQEQNKENLNPNSDAPVBS
KTSAPDYLVPAPPADIKTNRIYRIELNGYDPAYAGVFLGAKKRFGRKRTIWLFG
TKSAPDYLVPAPPADIKTNRIYRIELNGYDPAYAGVFLGAKKRFGRKRTIWLFG
ILGSKRVNDQAKSSAQIDPTPTVITSNTMCAVIDGNSITTFHQOPLQDRMFKPRL
TRLEHDFGKVTQEKVEPRFQADHVEVAHFYVRKGGANKRPAAPDDADKBPKEA
CPVADPSSDAEGAPVDPADRYQNKSRHAGMLQFLPCKTCTERNQNFNICTHGT
RDCSECFPGVSESQPVVRKRTYRKLCAIHLLGRAPEIACSDLVNVDLDDCVSEQ"
2208..4418
/codon_start=1
/evidence=not experimental
/product="capsid protein vp1"
/protein_id="AAB95450.1"
/db_xref="GI:2766607"
/translations="MAADGYLPDWLENDLSEIRWMDLKPAGPKPKANQOKQDDGRG
LVLPYKYLDPFNGLDGEPVNAADAALAEHDKAYDQOLKAGANPYLYRNHADAEPQ
RLQEDTSFGNLGRVFAKRLVLEPGLVEGAKTAPGKRPVQEPQSPSSGIG
DQGPAPKRLNFGTDSVPDPOPLGPPATPAAGVTWASGGGAPWADNNEGA
KTVGNAGNWHCDSTWLGDRVITTSRTWALPTVNNHLKQISSASTGASNDNHYFGY
STWGFDFNRFHCHSPRWQRLINNNGFRPKLNFKNQVKEVTNDGVTTIA
NNLTSTVQFSDSEYQLYVLSAHQCLPPFPADVFMIPOYGYLTLLNGSQVGRSS
FYCLEYPPSQMLRTGNNFTSYTFDVPFSSHSQSLDRLMPLIDQYLYLNRTQ
NQSAGAKNDLLFRSGSPAGVQKRWLPQCVQRQVSKTKTDNNNSNTWTGASK
YNLNGRESIIPGTMASHKDKQKPPMSGVMIFGKESASNTALDNVMTDEEEI
KATNPATERFGTVAVNLQSSSTDPTGTVHVMGALPGMWODRDVLTQYIWKAPIH
TDGHFPLPLMGFGGLKHPPLQILIKNTVPNPAPESATKFASTFOYSTGVSVSE
IWEMLQKENSQRNPEVQYTSNTAKSANDVFTVDNNGLYTEPRFGRILYLRPL"
BASE COUNT 1114 a 1363 c 1277 g 929 t
ORIGIN
Alignment Scores:
Pred. No.: 2,04e-229 Length: 4683
Score: 3229.00 Matches: 594
Percent Similarity: 99.67% Conservative: 3
Best Local Similarity: 99.17% Mismatches: 2
Query Match: 99.32% Indels: 0
DB: Gaps: 14
US-09-807-802A-15 (1-599) x AF028704 (1-4683)
Qy 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAenSerSerSer 20
Db 2619 ACGGCTCTGGAAGAAACGTCGCGTAGAGCAGTCGCCACAAAGAGCCAGACTCTCTCG 2678
Qy 21 GlyTleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAenPheGlyGlnThrGly 40
Db 2679 GGCATTGGCAAGACAGCCAGCAGCCGCTAAAAGAGACTCAATTTTGGTCAGACTGGC 2739
Qy 41 AspSerGluSerValProAenProGlnProLeuGlyGluProProAlaThrProAlaAla 60
Db 2739 GACTCAGAGTCAGTCCCGACCCCAACCTCTCGAGNACCTCCAGCNACCCCGCTGCT 2798
Qy 61 ValGlyProThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAenAenAenGlyGly 80
Db 2799 GTGGACCTACTACATGGCTTCAGCGGTCGCGCACCAATGGCAGACAATAAGAGGC 2858
Qy 81 AlaAspGlyValGlyAenAlaSerGlyAenTrpHisCysAspSerThrTrpLeuGlyAsp 100
Db 2859 GCCGACGAGTGGGTAAATGCTTCAGGAATTTGGCATTTGCCATGCTGGCGCAC 2918
Qy 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAenHisLeuTyr 120
```

```
Db 2919 AGAGTCATCACACACAGACCGACATGGGCTTGGCCACCTATACCAACCACTCTAC 2978
Qy 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer 140
Db 2979 AAGCAATCTCCAGTGTCTTCAACGGGGCGGACGACCAACCACTACTCTGGCTTACAGC 3038
Qy 141 ThrProTTPGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
Db 3039 ACCCCCTGGGGTATTTTGAATTCACAGATTCCACTGCCATTTCTCACACAGTGACTGG 3098
Qy 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLysLeuAsnPheLysLeuPhe 180
Db 3099 CAGCGACTCATCAACAACAAATTTGGGGATTCCGGCCCAAGAGACTCAACTTCAAGCTCTTC 3158
Qy 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeu 200
Db 3159 AACATCCAGTCAAGAGGTACAGACGAATGATGGCGTCACGACCATCGCTAATAACCTT 3218
Qy 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
Db 3219 ACCAGCAGGTTCAAGTCTTCTCGGACTCGGAGTACCAAGTTCGGCTGCTCTCGGCTCT 3278
Qy 221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGly 240
Db 3279 GCGCACACAGGCTGCTCCCTCCGTTCCGGCGGACGTTTCATGATTCGGCAGTACGCG 3338
Qy 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValAlaGlyArgSerSerPheTyrCysLeuGlu 260
Db 3339 TACCTAACGCTCAACAATGGCAGCAGGCGAGTGGGACGCTATCTTTTACTGCGCTGAA 3398
Qy 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
Db 3399 TATTTCCCATCGAGATCTGAGAACGGCAATTAATTTTACCTTCAGCTACACCTTCGAG 3458
Qy 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
Db 3459 GAGTGGCTTTCCACAGCAGTACGCGCACGACGAGCGCTGGACCGGCTGATGAATCT 3518
Qy 301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerIleGln 320
Db 3519 CTCATCGACGAGTACCTGATTTACCTGAAACAGAACTCAGAACTCAGTCGGGAAGTCCCAA 3578
Qy 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
Db 3579 AACAGGACTTGTCTGTTAGCCGGGCTCTCCAGCTGGCATGCTGTTTACGCCCCCAAAAC 3638
Qy 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
Db 3639 TGGCTACCTGGACCTGTTTACCGGACGAGCGCGCTTCTTAAACAAAAACAGAACCAAC 3698
Qy 361 AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 380
Db 3699 AACAGAACCTTTTACCTGGGACTGCTGCTTCAAAATATAACCTTAATGGGCGTGAATCTA 3758
Qy 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet 400
Db 3759 ATCAACCTTGGCACTGCTATGGCTCACAAAGACGACCAAGACAACTTCTTTCCCATG 3818
Qy 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
Db 3819 AGCGGTGTATGATTTTGGAAAGGAGAGCGCGGAGCTTCAACACACTGCATTGGACAAT 3878
Qy 421 ValMetIleThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
Db 3879 GTCATGATCACAGCAAGAGGAAATCAAAAGCCACTAAACCCGCTGGCCACCGAAGATTT 3938
Qy 441 GlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHis 460
Db 3939 GGGACTGTGGCACTCAATCTCCAGAGCAGCAGCAGACGACCTGCGACCGGAGATGTCAT 3998
Qy 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro 480
```

```
Db 3999 GTTATGGAGCGCTTACCTGGAATGGTGTGGCAAGACAGACAGCTATACCTGCAGGGTCT 4058
Qy 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
Db 4059 ATTTGGGCCAAAATTCCTCACAGGATGGACACTTTTCCACCGCTCTCTCTCATGGGCGC 4118
Qy 501 PheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsn 520
Db 4119 TTTGAGCTTAAGCACCCGCTCTCAGATCTCATCAAAACACGCTGTTCTCTCGCAAT 4178
Qy 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
Db 4179 CCTCCGCGAGATTTTCCGCTACAAAGTTTGTCTTTCATTCATCCCAAGTATTCACAGGA 4238
Qy 541 GlnValSerValGluIleGluTyrGluLeuGlnLysGluAsnSerLysArgTyrAsnPro 560
Db 4239 CAAGTGAGCGTGGAGATTGAATGGAGCTGCAGAAAGAAACAGCAACGCTGGAATCCC 4298
Qy 561 GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsn 580
Db 4299 GAAGTGCAGTATACATCTAATGCAAAATCTGCCAACGTTGATTTTCACTGTGGCAAC 4358
Qy 581 AsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db 4359 AATGAGCTTTTATAGCTGAGCTCGCCCAATGGCAACCGTTACCTCACCCGCTCCCTG 4415

RESULT 3
AF028705 4722 bp DNA linear VRL 12-JAN-1998
LOCUS Adeno-associated virus 3B, complete genome.
DEFINITION AF028705
ACCESSION AF028705.1 GI:2766608
VERSION
KEYWORDS Adeno-associated virus 3B
SOURCE Adeno-associated virus 3B
ORGANISM Adeno-associated virus 3B
REFERENCE 1 (bases 1 to 4722)
AUTHORS Rutledge,E.A., Halbert,C.L. and Russell,D.W.
TITLE Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2
JOURNAL J. Virol. 72 (1), 309-319. (1998)
MEDLINE 98080418
PUBMED 9420229
REFERENCE 2 (bases 1 to 4722)
AUTHORS Rutledge,E.A. and Russell,D.W.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-1997) Hematology, Univ of Washington, Box 357720, Seattle, WA 98195, USA
FEATURES
source
1. 4722
/organism="Adeno-associated virus 3B"
/mol_type="genomic DNA"
/db_xref="taxon:68742"
/notes="AAV3B"
317..2191
/codon_start=1
/product="nonstructural protein Rep78"
/protein_id="AAB95451.1"
/db_xref="GI:2766609"
/translation="MPGFYEIVLVKPSDLDEHLPGISNSFVNVAEKEMELPPDSMD
PNLEQAPLTVAEKLEFLVWRVSKAPEALFFVQFEKGETYFHLHLVLETIGVKS
MVVGRYSQIQKEKLVTRIYRGVBPQPNFVATKTRNGAGGKGVDDCYIPNYLLPK
TQPELQAWNTNDQYLSACLNAERKRLVAQHLTHVSQTQBNKQENQPNQDAVIRS
KTSARYMELVGMVDRGITSKQWIOEDQASYISFNAASNSRQIKAAALNASKIMSL
KTADPYLVGNSNPEDITKNRIYQILELNGDPQVAAASVPLGWAQKFKQKNTIWLFG
PATGKTNIARAIHAHVPYGCYVNTNENFPNDCYDQWVIMWEEKTKAKVSESALA
ILGSKVRVDQKSSAQIBFPIVITSNTNNCAVIDGNTSTTFHQQLQDRMPFEL
TRRLDHPGKVTQSEVDQPFKMSADHDVDAFVYRGKGAKKRSPASNDVSEPKRQ
CTSLAQPTTSDAEPADYADYQNKSRHVGKMLFPFKCTCERNQISNVCFTHGQR
DCGECFPGMSSESQPVSVVKKTYQKLCPIHHILGRAPEIACSDLANVLDLDCVSEQ
"
2208..4418
/codon_start=1
```



```

/product="capsid protein VP1"
/protein id="AAB95452.1"
/db_xref="GI:2766610"
/translation="MAADGVLPMLENDLSEIREWALKPGVPQPKANQOHOHNRG
LVLPYKYLPGNGLDGEFVNEADAALAHEDKAYDQLKAGNDPYLYKXHADAPFG
RLQEDTSFGNIGRAVFOAKRIEPLGLVEEAATAPGKRKPVDPQSPDSGGVG
KSGKOPARKLNFQOTGDSSEVPDPQLGEPFAAPTSGNTWASGGGAPMADNNEGA
DGVNSSGNHCDQMWLDGDRVITTSRTWALPTNNHLYKQISSQSGASNDNHVFGYS
TPMGYFDFNRCHFSRPMORLNNNNRPKPLKFLNIQVKEVYONDGTITIAN
NLTSVQVFTDSEVOLPVLVLSAQHGLPPFPADVPOYGVYLTLANGSOAVGRSSF
YCLRYFSPOMLRTGNPFSTYFEDVPFHSYAHQSILDLMLPLIQYLYLNRTOG
TTSQTNQSLRFGDQSGMSLQARNLPGFCRTQRQRLSKTANDNNNSFPMTAASK
YHLNGLSLNPGFAMASHKDDDEKFPFMHGNLIPGKEGTTASNAEDLNMWITDEBI
RTNPTVATEQYGTVANLLSQNTAPTRTNDQGLFCMVWQDRDVLQGPFIWAKIPIH
TDGHFSPPLMGGLGHPPIQIMIKNTVPANPPTTSPAKFASFTQYSTGVSVSE
IEWELOSNSRKNPEIQYTSYNNKSNVNDFTVDNGVYSEPRPIGTYLRLNL"
BASE COUNT 1204 a 1297 c 1243 g 978 t
ORIGIN

Alignment Scores:
Pred. No.: 5,47e-200 Length: 4722
Score: 2832.00 Matches: 514
Percent Similarity: 91.33% Conservative: 34
Best Local Similarity: 85.67% Mismatches: 50
Query Match: 87.11% Indels: 2
DB: 14 Gaps: 2

US-09-807-802A-15 (1-599) x AF028705 (1-4722)

Qy 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSer 20
Db 2619 ACGGCTCTGGAAAGAACAGCGCTGTAGTCAGTCTCTCCAGGACCGGACTCATCTCT 2678

Qy 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLysLeuAsnPhelGlnThrGly 40
Db 2679 GGTGTGTGCAAAATCGGGCAACAGCGCTGCCAGAAAAGACTAAATTCGGTCAGACTGGC 2738

Qy 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
Db 2739 GACTCAGAGTCAGTCCAGACCCCTCAACCTCTCGGAGAACACACAGCAGCCCCACCAAGT 2798

Qy 61 ValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGly 80
Db 2799 TGGGATCTAATACAAATGGCTTCAGCGGTGGCGCACCAATGGCAGACAAATAACGAGGGT 2858

Qy 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTriLeuGlyAsp 100
Db 2859 GCCGATGGAGTGGGTAAATCTCAGGAATTTGGCAATTCGGATTCCCAATGGCTGGCGAC 2918

Qy 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrThrAsnAsnHisLeuTyr 120
Db 2919 AGAGTCATCACACCAGCAGCAACCTTGGGCGCTCCGCACTTCAACCAACCACTCTCTAC 2978

Qy 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer 140
Db 2979 AAGCAAAATCTCCAGCAATCA--GGAGCTTCAACAGCAACCACTACTTTGGCTACAGC 3035

Qy 141 ThrProTTPGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
Db 3036 ACCCTTGGGGTATTTTGACTTTTAAACAGATTCACATCTCCACTTCTCACACGTCGACTGG 3095

Qy 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPhelLysLeuPhe 180
Db 3096 CAGCGACTCATTAACAACAACTGGGGATTCGGGCCCAAGAACTCAGCTTCAAGCTCTTC 3155

Qy 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeu 200
Db 3156 AACATCCAAAGTTAAAGAGGTACGCGAGAACGATGGCACGAGCTATTGGCCAAATACCTT 3215

Qy 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
Db 3216 ACCAGCAGGTTCAGTGTTCACGACTTCGGAGTATCAGTCCCGTACGTCTCGGGTCG 3275

```

```

221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGly 240
3276 GCGACCAACAGGCTGTCTCCGCGCTTCCAGCGGAGCGTCTTATGTCCTCAGTAGTGA 3335

Qy 241 TyrIleuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
3336 TACTTCACCTCGAACAACGGAAGTCAAGCGGTGGAGCGCTCATCTTTTAACTGCTCGAG 3395

Qy 261 TyrPheProSerGlnMetLeuAArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
3396 TACTTCCTTCGAGAGTCTAAGAGCTGGAAATAACTTCCAAATTCAGCTATACCTTCGAG 3455

Qy 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
3456 GATGTACTCTTTCACAGCAGCTACGCTCACAGCAGAGTTTGGATCGCTTGTGATGAATCCT 3515

Qy 301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGln--AsnGlnSerGlySerAla 319
3516 CTTATTGATCAGTATCTGTACTTCTTGAACGAGACGCAACCACTCTGGACCAACC 3575

Qy 320 GlnAsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLys 339
3576 AACCAATCACGGCTGCTTTTTCAGCAGGTGGCGCTCAGTCTATGCTCTTTCGAGGCCAGA 3635

Qy 340 AsnTrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsn 359
3636 AATTGGCTACTCGGGCCCTGCTACCGGCAACAGAGACTTCAAGAGACTGCTTAACGACAAAC 3695

Qy 360 AsnAsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSer 379
3696 AACCAAGTAACTTCTTGGACAGCGGCCGCAATATCATCTCAATGTCGCGGACTCG 3755

Qy 380 IleIleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePhePro 399
3756 CTGTGTAATCCAGACCGACTATGGCCAGTCACAAGGACGATGAAGAAATAATTTTCCCT 3815

Qy 400 MetSerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAsp 419
3816 ATGCAGCGCAATCTAATATTGGCAAGAGGACCAACGCAAGTAAACGCAAGATTAAGT 3875

Qy 420 AsnValMetIleThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArg 439
3876 AATGTAATGATTACGATGAGAGAGATTCGTACCAACCAATCTCTGTGGCAACAGAGCAG 3935

Qy 440 PheGlyThrValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspVal 459
3936 TATGGAATCTGGCAATAACTTCAGAGCTCAAAATACAGCTCCACGACTAGAACTGTCTC 3995

Qy 460 HisAlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGly 479
3996 AATGATCAGGGGCGCTTACCTGGCATGTGTGGCAAGATCGTACGCTGTACCTTCAAGGA 4055

Qy 480 ProIleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGly 499
4056 CCTATCTGGCAAGATTCCTCACCGATGGACACTTTCATCTCTCTCTCTCTGATGGA 4115

Qy 500 GlyPheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAla 519
4116 GGCTTTGGACTGAACAACATCGCTCTCTCAATCATGATCAAAATACTCCGGTACCGGCA 4175

Qy 520 AsnProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThr 539
4176 AATCTCTCCGACGACTTTCAGCGCGGCAAGTTTGTCTTATTCATCTCAGTACTCCACT 4235

Qy 540 GlyGlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsn 559
4236 GGCAGGTCTAGCTGGAAATTTGATGGGAGCTACAGAAAGAAACAGCAACGTTGGAAT 4295

Qy 560 ProGluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAsp 579
4296 CCAGAGATTCTAGTACACTTCCAACTACAAACAGTCTGTTAATGTGGAGCTTTTACTGTAGAC 4355

Qy 580 AsnAsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599

```

DB 4356 ACTAATGGTGTATAGTGAACCTCGCCTATTGGAACCGGTATCTCACACGAACTTG 4415

RESULT 4
AVU48704 4726 bp DNA linear VRL 15-JUL-1996
LOCUS Adeno-associated virus 3 nonstructural protein and capsid protein
DEFINITION genes, complete cds, and complete genome.
ACCESSION U48704
VERSION U48704.1 GI:1408467

KEYWORDS
SOURCE
ORGANISM Adeno-associated virus 3
Adeno-associated virus 3
Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
REFERENCE 1 (bases 1 to 4726)
Muramatsu, S., Mizukami, H., Young, N.S. and Brown, K.E.
Nucleotide sequencing and generation of an infectious clone of
adeno-associated virus 3
TITLE
Virology 221 (1), 208-217 (1996)

JOURNAL
MEDLINE 96286430
PUBMED 8661429
REFERENCE 2 (bases 1 to 4726)
Muramatsu, S. and Brown, K.E.
AUTHORS
Direct Submission
TITLE
Submitted (06-FEB-1996) Shin-ichi Muramatsu, Hematology Branch,
NHLBI/NIH, 9000 Rockville, MD 20892, USA
JOURNAL
Location/Qualifiers

FEATURES
source
1..4726
/organism="Adeno-associated virus 3"
/mol_type="genomic DNA"
/strain="3H"
/db_xref="taxon:46350"
318..2192
/codon_start=1
/product="nonstructural protein"
/protein_id="AAC5048.1"
/db_xref="GI:1408468"
/translation="MRGFYELVLPSPDLRLPGISFNWVAEKWDVPPDSMD
PNLQYQVTAELKQRELFVWRVSKAPEALFFVFEKSGTYFHLVLVETIGVKS
MVGRYSQIKLELVIRYGEVQPLPFWNFATVTRNGAGGKGVDDCIPYLLLPK
TOPELQWANTMDVLSACLAEKRLVAOHLTVSOTQONKQNPNSDAPVRS
KTSARWELVMDVDRGTSIKQVDSQYISFNAASRSQIKALDNASKMISL
TKTAPLVGSPPEPDKTNRVYQILELNGDPOVAASVFLGWAOKFKRNTIWLFG
PATYKNTAEALAHAPVPGVWNTNENFPENDCVDMVIMWESGKWTAKVESAKA
ILGSKVRVDQKCSAQIEPTPIVTSNTNCAVIDGNTTFEHOQPLQDRMFEEL
TRLDRHFGKTVQEVDFRWDHVDVHEFYVRGGAKKRPSNDADVSEPRE
CTSLAQPTTSDAEAPADYADYONKCSRHVGNMLFPCKTCERNQISNVCFTHGQR
DCGECFPGMSQPSVSWVKTKYQKLCPIHILGRAPEIACSAACDLANVDLDDCVSEQ
"

CDS
2209..4419
/codon_start=1
/product="capsid protein"
/protein_id="AAC5049.1"
/db_xref="GI:1408469"
/translation="MAADGYLPDLENDLSEGIWVWALPGVPOPKAQQHODNRGR
LVLPGYLPNGGLGKQVNEADAALHDKAYDQQLKAGNPVLYKNHDAADFQ
RLQEDTSGNGLRVAQKRLLEPLGLVEAKTAPKKGAVDQSPPEPSSSGVG
KSGKPARLKNLPQGDSEVPDQPLGEPPAATSLGNTWASGGGAPMADNNEGA
DVGNSGNWHCDQSLDRVITSTRWALPTNNHLYKQISSGASNDNHYFGYS
TPWGFDPFRFCHFSQPRDWRLINNWGFPRPKLSFKLFNIQVRGTQNDGTIAN
NLSTVQVTPSYQLPYVLGSAHQCLPPFPADVPWQYGYLLNNGSAVGRSSF
XCLYEPFQMLTGNFQSYTFEDVPFHSVYAHOSLDRLMPLIDQYLYLNRTQG
YCTSGTNTQSLKLSQAGQMSQARNLPGPCYRQRLSKTDNNDNNSFPWFTASK
YHLNGRDSLVPNGPASHKDDBEKFFPMHNGLIFGEGTATASNLNDNMVLTDEBI
RTTNVPATGYGVANNQSSNTAPTTGTVNHQGLPQWQDRDYLQGLPWAKLIPH
TDGTFPSPGLMGFGKLPQPIQMIKNTVPANPPTTFSPAKFASITQYSTQVSVSE
IWELOKENSKNWPEIQVTSYNTKSNVNDFTVDVTNGVSEPRPITGRYLRNL"

BASE COUNT 1200 a 1295 c 1251 g 980 t
ORIGIN
Alignment Scores:
Pred. No.: 9.97e-199 Length: 4726
Score: 2815.00 Matches: 511

Percent Similarity: 91.00% Conservative: 35
Best Local Similarity: 85.17% Mismatches: 52
Query Match: 86.59% Indels: 2
DB: Gaps: 2

US-09-807-802A-15 (1-599) x AVU48704 (1-4726)
Qy 1 ThrAlaProGlyValValArgProValGluGlnSerProGlnGluProAspSerSerSer 20
Db 2620 ACGGCTCTCTGGANAGAGGGGGCTGTAGATCAGTCTCTCTCAGGAAACCGGACTCATCT 2679
Qy 21 GlyIleGlyValThrGlyGlnGlnProAlaIleValArgLeuAenPheGlyGlnThrGly 40
Db 2680 GGTGTTGGCAAAATCGGGCAAAACAGCTTCCAGAAAAAGACTTAAATTTTCGGTCCAGATCGA 2739
Qy 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
Db 2740 GACTCAGAGTCAGTCCAGAGACCTTCAACCTCTCCGAGAACACCAGCAGAGCCCCCAGCAGT 2799
Qy 61 ValGlyProThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGluGly 80
Db 2800 TTGGGATCTAATCAATGGCTTTCAGGCGGTGGCCACCAATGGCAGACCAATAACGAGGGT 2859
Qy 81 AlaAspGlyValGlyValAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
Db 2860 GCCGATGGAGTGGGTAAATTCCTCAGGAAATTTGGCATTTCCGATTTCCCAATGCTGGCGCAC 2919
Qy 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTrpAsnAsnHisLeuTyr 120
Db 2920 AGAGTCATCACCACGACGACCAACCTGGGCGCTCCGCCACTTACCAACACCAATCTCTAC 2979
Qy 121 LysGlnIleSerSerAlaSerThrGlyValSerAsnAspAsnHisTyrPheGlyTyrSer 140
Db 2980 AAGCAATCTCCAGCCCAATCA--GGAGCTTCAACGACACCACTACTTTGGCTACAGC 3036
Qy 141 ThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
Db 3037 ACCCTTGGGGTATTTTACATTTTAAACAGATTCCACTGCCACTTCTCACCACGTCGACTGG 3096
Qy 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheValLeuPhe 180
Db 3097 CAGCGACTCANTAAACAACAACTGGGGATTCGGGCCCAAGAAACCTCAGCTTCAAGCTCTTC 3156
Qy 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeu 200
Db 3157 AACATCCAGTTAGAGGGGTCCAGCAGAACGATGCGCAGCAGACTATTGGCCAATACCTT 3216
Qy 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
Db 3217 ACCAGCAGGTTCAGAGTGTTCAGGACTCGGAGTATCAGCTCCCGTACGTGCTCGGGTCG 3276
Qy 221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGly 240
Db 3277 CGCGACCAAGGCTGTCTCCCGCGTTTCCAGCGGAGCTTTCATGCTGCTCCTCAGTATGGA 3336
Qy 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
Db 3337 TACCTCACCTGAACAACGGAAGTCAAGCGGTGGGACGCTCATCTTTTACTGCTGGAG 3396
Qy 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
Db 3397 TACTTCCCTTCGAGATGCTAAGGACTGGAATAAATCTTCCAATTCAGCTATATACCTTCGAG 3456
Qy 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
Db 3457 GATGTACCTTTTTCAGCAGACGTACCTCACAGCAGAGTTGGATCGCTTGATGATCTCT 3516
Qy 301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGln---AsnGlnSerGlySerAla 319
Db 3517 CTTATTGATCAGTATCTGTACTTACCTGAACAGAACGAAAGGAAACACCTCTGGAACACCC 3576
Qy 320 GlnAsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLys 339

Db 3441 CTTATTGTAGTACATCTGTACTACCTGAACAGAACGCIAGGAACAACCTCTGGAACAACC 3500
Qy 320 GlnAsnLysAspLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLys 339
Db 3501 AACCAATACAGCGCTGCTTTTATGACGAGCTGGCGCTCAGTCTATGCTTTTGCAGGCCAGA 3560
Qy 340 AsnTrpLeuProGlyProCysTrpArgGlnGlnArgValSerLysThrAspAsn 359
Db 3561 AATGGCTACCTGGGCGCTGCTACCGGCAACAGAGACTTTCAGAGACTGCTTAACGACAAAC 3620
Qy 360 AsnAsnSerAsnPheThrTrpThrGlyAlaSerLysTrpAsnLeuAsnGlyArgGluSer 379
Db 3621 AACACAGTAATCTTCTTGGACGCGCCAGCAATATATCTCATGGCGCGCACTCG 3680
Qy 380 IleLeuAsnProGlyThrAlaMetAlaSerHisLysAspGluAspLysPhePro 399
Db 3681 CTGTGTAATCCAGGACCACTATGCGCAGTCAAGAGGACGATGAAGAAAAATTTTCCCT 3740
Qy 400 MetSerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAsp 419
Db 3741 ATGCACGGCAATCTAATATTTGGCAAGAGGACCAACGGCAAGTAACGAGAAATTAGAT 3800
Qy 420 AsnValMetIleThrAspGluGluLysAlaThrAsnProValAlaThrGluArg 439
Db 3801 NATGTAATGATACGGATGAAGAGAGATTCTGACCACTATCTCTGTGGCAACAGAGCAG 3860
Qy 440 PheGlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspVal 459
Db 3861 TATGGAACCTGTGGCAAAATTAATTTGCAGAGCTCAAAATACAGCTCCCAACAGCTGGAATGTC 3920
Qy 460 HisAlaMetGlyValAlaLeuProGlyMetValTrpGlnAspArgAspValTrpLeuGlnGly 479
Db 3921 AATCATCAGGGGGCTTACTCGCATGGTGTGGCAAGATCGTGAGCTGATGATCATCTTCAAGA 3980
Qy 480 ProfileTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGly 499
Db 3981 CCTATCTGGGCAAGATTCCTCACAGGATGGACACTTTTCATCTCTCTGTATGGA 4040
Qy 500 GlyPheGlyLeuLysAsnProProGlnIleLeuLysAsnThrProValProAla 519
Db 4041 GGCCTTTGGACTGAACATCCGCTCTCTCAATCATGATCAAAATATCTCCGCTACCTGG 4100
Qy 520 AsnProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTrpSerThr 539
Db 4101 AATCCTTCAGCACCTTACGTCGGGCAAGATTGCTTCTTCTATCACAGTACTCCACG 4160
Qy 540 GlyGlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsn 559
Db 4161 GGACAGGTCAAGCTGGAGATCGAGTGGAGCTGCAGAGGAAACAGCAACGCTGGAAT 4220
Qy 560 ProGluValGlnTrpThrSerAsnTrpAlaLysSerAlaAsnValAspPheThrValAsp 579
Db 4221 CCGGAATTCAGTACACTTCCCACTACAAAGTCTGTTAATCGTGGACTTACCGTGGAT 4280
Qy 580 AsnAsnGlyLeuTrpThrGluProArgProIleGlyThrArgTrpLeuThrArgProLeu 599
Db 4281 ACTAATGGGTGATTTACAGAGCTCGCCCAATTTGGCACCAGATACCTGACTGTAATCTG 4340

RESULT 6

AR222046 LOCUS 4072 bp DNA linear PAT 26-SEP-2002
Sequence 4 from patent US 6429001.

AR222046

AR222046

AR222046.1 GI:23329402

Unknown.

Unknown.

Unclassified.

1 (bases 1 to 4072)

Hardy, S.F.

Recombinant AAV packaging systems

Patent: US 6429001-A 4 06-AUG-2002;

Location/Qualifiers

FEATURES

source 1..4072
BASE COUNT 1073 a 1025 c 992 g 982 t
ORIGIN
Alignment Scores:
Pred. No.: 4,51e-195 Length: 4072
Score: 2764.50 Matches: 496
Percent Similarity: 89.98% Conservative: 43
Best Local Similarity: 82.80% Mismatches: 59
Query Match: 85.04% Indels: 1
Gaps: 6
US-09-807-802A-15 (1-599) x AR222046 (1-4072)
Qy 1 ThrAlaProGlyLysLysArgProValGlnSerProGlnGluProAspSerSerSer 20
Db 1895 ACGGCTCGGGAAAAAGAGCGCGGTAGAGCACTCTCTGTGGAGCCAGACTCTCTCTCG 1954
Qy 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGly 40
Db 1955 GGAACCGGAAAGCGCGCCAGCAGCTGCAAGAAAAAGATTGGAATTTTGGTCAGACTGGA 2014
Qy 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
Db 2015 GACGAGACTCAGTACTGACCCCGCAGCTCTCGGACAGCCACCAGCAGCCCCCTCTG 2074
Qy 61 ValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGly 80
Db 2075 CTGGGAATAATAATCGATGGCTTACAGGCGAGTGGCGACCAATGGCAGACAAATAACGAGG 2134
Qy 81 AlaAspGlyValGlyValAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
Db 2135 GCCACGAGTGGGTAAATTCCTCGGAAATTTGGCATTCACATGATGGCGGCAC 2194
Qy 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTrpAsnAsnHisLeuTrp 120
Db 2195 AGAGTCATCACCAACGACCCCGAACCTGGGCGCTGCCACCTCAACAACACCCCTCTAC 2254
Qy 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTrpPheGlyTrpSer 140
Db 2255 AAACAAATTTCCAGCCCAATCA---GGAGCTTCGAAACGACAAATCACTACTTTGGCTACAG 2311
Qy 141 ThrProTrpGlyTrpPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
Db 2312 ACCCTTGGGGGTATTTGACTTCAACAGATTCCACTGCCACTTTTACCAGCTGACTGG 2371
Qy 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
Db 2372 CAAGACTCATCAACAACACTGGGGATTCCGACCCCAAGAGACTCAACTTCAAGCTCTTT 2431
Qy 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeu 200
Db 2432 AACATTCAGTCAAAAGAGGTCAACAGAAATGACGGTACGACGACGATTTGCCAATAACCTT 2491
Qy 201 ThrSerThrValGlnValPheSerAspSerGluTrpGlnLeuProTrpValLeuGlySer 220
Db 2492 ACCAGACGGTTTACGGTGTCTTACTGACTCGGAGTACCAAGCTCCCGTACGCTCTCGGCTCG 2551
Qy 221 AlaHisGlnGlyCysLeuProPheProAlaAspValPheMetIleProGlnTrpGly 240
Db 2552 GCGCATCAAGGATGCTCTCCCGCGCTTCCAGCAGACAGCTCTTTCATGGTCCACAGTATGGA 2611
Qy 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTrpCysLeuGlu 260
Db 2612 TACCTCACCTCTGAACACACGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCA 2671
Qy 261 TyrPheProSerGlnMetLeuArgThrGlyValAsnAsnPheThrPheSerTrpThrPheGlu 280
Db 2672 TACTTTCTCTCAGATGCTGGTACCGGAACCACTTTTACCTTCAGCTACACTTTTGG 2731
Qy 281 GluValProPheHisSerSerTrpAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
Db 300

Db 2732 GACGTCCTTCCACAGCAGCTACGCTCAGACGAGAGCTGGACCGCTCTCATGAATCCT 2791
Qy 301 LeuileAepGlnTyrLeuTyrTyrLeuAenArgThrGlnAenGlnSerGlySerAlaGln 320
Db 2792 CTCTCGACCACTACCTGTATTTCTTGAGCAGACAAACACACTCCAGTGGACACCACG 2851
Qy 321 AsnLysAepLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAen 340
Db 2852 CAGTCAAGCCTTCACTTCTCAGCGCCGAGCGAGTGACATTCGGGACCACTAGGAAC 2911
Qy 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAen 360
Db 2912 TGGCTTCTCGACCTGTGTACCGCAGCGAGGATCAAGACATTCGCGGATAACAC 2971
Qy 361 AsnSerAenPheThrTrpThrGlyAlaSerLysTyrAenLeuAenGlyArgGluSerIle 380
Db 2972 AACAGTGAATACTCGTGACTGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTG 3031
Qy 381 IleAenProGlyThrAlaMetAlaSerHisLysAepAepGluAepLysPhePheProMet 400
Db 3032 GTGAATCCGGGCCCGCCATGCGAAGCCACAGGACGATGAAGAAAGTGTTCCTCAG 3091
Qy 401 SerGlyValMetIlePheGlyValGluSerAlaGlyAlaSerAenThrAlaLeuAenAen 420
Db 3092 AGCGGGTCTCTATCTTTGGGAGAGCGCTCAGAGAAACAAATGTGGACATTTGAAG 3151
Qy 421 ValMetIleThrAepGluGluLysAlaThrAenProValAlaThrGluArgPhe 440
Db 3152 GTCATGATACAGACGAGAGGAAATCAGGACCAACCAATCCGTTGGCTACGAGCAGTAT 3211
Qy 441 GlyThrValAlaValAenPheGlnSerSerSerThrAepProAlaThrGlyAepValHis 460
Db 3212 GGTTCGTATCTTACCAACTCTCCAGAGAGCAACAGACAGAGCTTACCGCAGATGTCAAC 3271
Qy 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAepArgAepValTyrLeuGlnPro 480
Db 3272 ACACAGCGGTTCTCCAGCGATGTCTGGCAGGACAGAGATGTATCCTTCAGGGGCC 3331
Qy 481 IleTrpAlaLysIleProHisThrAepGlyHisPheHisProSerProLeuMetGlyGly 500
Db 3332 ATCTGGCAAGATTTCCACACAGGACGAGACATTTTACCCCTCTCCCTCATGGTGA 3391
Qy 501 PheGlyLeuLysAenProProGlnIleLeuLysAenThrProValProAlaAen 520
Db 3392 TTCGAGATTAAACACCTCTCCACAGATTTCTCATCAAGAACACCCCGGTACCTGCGAAT 3451
Qy 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
Db 3452 CCTTCGACACCTTCAGTGGCGCAAGTTTGCTCTTCATCACACAGTATCCACGGGA 3511
Qy 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAenSerLysArgTrpAenPro 560
Db 3512 CAGGTCAAGTGGAGATCAGTGGGAGCTGCAGAGGAAACAGCAACGCTGGAATCCC 3571
Qy 561 GluValGlnTyrThrSerAenTyrAlaLysSerAlaAenValAepPheThrValAenAen 580
Db 3572 GAAATTCAGTACACTTCCAACTACAAACAGTCTGTATATGTGGACTTTACTGTGGACACT 3631
Qy 581 AenGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db 3632 AATGGCGGTATTACAGAGCTCGCCCATTTGGCACCAGATACCTGACTCGTAATCTG 3688

RESULT 7
AX205074 4072 bp DNA linear PAT 30-AUG-2001
LOCUS Sequence 4 from Patent WO0155361.
DEFINITION AX205074.1
ACCESSION AX205074.1 GI:15394315
VERSION
KEYWORDS
SOURCE
ORGANISM
unidentified
unclassified.
REFERENCE 1

AUTHORS Hardy, S.P.
TITLE Recombinant aav packaging systems
JOURNAL Patent: WO 015361-A 4 02-AUG-2001;
Chiron Corporation (US)
FEATURES Location/Qualifiers
source 1..4072
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/notes="recombinant DNA"

BASE COUNT 1073 a 1025 c 992 g 982 t
ORIGIN

Alignment Scores: 4.51e-195 Length: 4072
Pred. No.: 2764.50 Matches: 496
Score: 2764.50
Percent Similarity: 89.98% Conservative: 43
Best Local Similarity: 82.80% Mismatches: 59
Query Match: 85.04% Indels: 1
DB: 6 Gaps: 1

US-09-807-802A-15 (1-599) x AX205074 (1-4072)

Qy 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAaspSerSer 20
Db 1895 ACGGCTCCGGGAAAAAGAGCGCGGTAGACACTCTCTCTGGAGCCAGACTCTCTCTCG 1954
Qy 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAenPheGlyGlnThrGly 40
Db 1955 GGAACCGAAAGCGGGCCAGCAGCTCGCAAGAAAAAGATTGAATTTGTTCAGACTGGA 2014
Qy 41 AspSerGluSerValProAepProGlnProLeuGlyGluProProAlaThrProAlaAla 60
Db 2015 GAGCAGACTCAGTACCTGACCCCGAGCTCTCGACACACACAGCCCTCTCGGT 2074
Qy 61 ValGlyProThrThrMetAlaSerGlyGlyValaProMetAlaAepAenAenGluGly 80
Db 2075 CTGGGAACATAATAGATGGCTCAGGCGATGGCGCACCAATGGCAGACATNACGAGGC 2134
Qy 81 AlaAepGlyValGlyAenAlaSerGlyAenTrpHisCysAepSerThrTrpLeuGlyAep 100
Db 2135 GCCGACGAGTGGGTAAATCTCGGAAATTTGGCATTGGCATTCACATGGATGGCGAC 2194
Qy 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTyrAenAenHisLeuTyr 120
Db 2195 AGAGTCAATCACACACAGCAGCAGCCGAACTGGGCGCTGCCCCACCTACAAACACCACTAC 2254
Qy 121 LysGlnIleSerSerAlaSerThrGlyValaSerAenAepAenHisTyrPheGlyTyrSer 140
Db 2255 AAACAAATTTCCAGCCCAATCA---GGAGCTCGACGACATCACTACTTTGGCTACAGC 2311
Qy 141 ThrProTrpGlyTyrPheAepPheAenArgPheHisCysHisPheSerProArgAepTrp 160
Db 2312 ACCCTTTGGGGTATTTTGAATTCACAGATTCACCTGCCACTTTTCCACACGACTGCTG 2371
Qy 161 GlnArgLeuIleAenAenAenTrpGlyPheArgProLysArgLeuAenPheLysLeuPhe 180
Db 2372 CAAAGACTCATCAACAACTCGGGATTCGAGCCCAAGAGACTCAACTTCAGAGCTCTTT 2431
Qy 181 AsnIleGlnValLysGluValThrThrAenAepGlyValThrThrIleAlaAenAenLeu 200
Db 2432 AACATTCAAGTCAAGAGAGTCAAGCAGATGACGGTACGACGACGATTTGCCAATACTT 2491
Qy 201 ThrSerThrValGlnValPheSerAepSerGluTyrGlnLeuProTyrValLeuGlySer 220
Db 2492 ACCAGCAGGTTTCAAGTGTATTTACTGACTCGAGTACCAAGTCTCCGCTCGCTCGCTCG 2551
Qy 221 AlaHisGlnGlyCysLeuProProPheProAlaAepValPheMetIleProGlnTyrGly 240
Db 2552 GCGCATCAAGGATGCTCCCGCGCTTCCAGCAGACGCTCTTCATGGGGCCACATATGGA 2611
Qy 241 TyrLeuThrLeuAenAenGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260

```
Db 2612 TACCTCACCCCTGAACAACGGGAGTCAGGAGTACGAGCGCTCTTCATTTTACTGCTGGAG 2671
Qy 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerThrPheGlu 280
Db 2672 TACTTTCTCTCTCAGATGCTGCTACCGGAAACACACTTTTACCTTCAGCTACACTTTTGAG 2731
Qy 281 GluValProPheHisSerSerThrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
Db 2732 GACGTTCCTTTCCACAGCAGCTACCTTCACGCGAGAGTCTGGACCGTCTCATGATCT 2791
Qy 301 LeuIleAspGlnTyrLeuTyrThrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
Db 2792 CTCATCGACCACTACCTGATTTACTTGAGCAGAACAAACACTCCCAAGTGGAAACACACAG 2851
Qy 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
Db 2852 CAGTCAAGGCTTCAGTTTCTCAGCCCGGAGCGAGTGACATTCGGGACAGCTTAGGAAC 2911
Qy 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
Db 2912 TGCTTCTCTGGACCTTGTTTACCAGCAGCAGCAGATTCAGACATCTGCGGATACACAC 2971
Qy 361 AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgLeuSerIle 380
Db 2972 AACAGTGAATACCTCGTGGAGCTGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTG 3031
Qy 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPheProMet 400
Db 3032 GTGAATCCGGGCGCGCCCATGCAAGCCACAGGACGATGAAGAAAGTTTTCCTCAG 3091
Qy 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
Db 3092 AGCGGGTCTCATCTTTGGAGACAGAGCTCAGAGAAACAAATGTGGACATCGAAAG 3151
Qy 421 ValMetIleThrAspGluGluLeuLysAlaThrAsnProValAlaThrGluArgPhe 440
Db 3152 GTCATGATTACAGACAGAGGAAATCAGACAAACCAATCCGCTGGCTACGGAGCAGTAT 3211
Qy 441 GlyThrValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHis 460
Db 3212 GGTTCTGTATCTACCAACTCTCAGAGAGGACAGACAGAGCTACCCGATGTCAAC 3271
Qy 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro 480
Db 3272 ACACAAGGCGTCTTCCAGCATGGTCTGGCAGGACAGAGATGTGTACCTTCAGGGGCC 3331
Qy 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
Db 3332 ATCTGGGCAAGATTCCACACAGCAGGACGACATTTTCACCCCTCTCCCTCATGGGTGA 3391
Qy 501 PheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsn 520
Db 3392 TTCGGACTTAAACACCCCTCTCTCAGAGATCTCATCAAGAACACCCCGTACTCGGAAT 3451
Qy 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
Db 3452 CCTTCGACCACTTCAGTGGGCAAGTTTGTCTTCTTCATCACACAGTACTCCAGGGA 3511
Qy 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPro 560
Db 3512 CAGGTCAAGTGGAGATCGAGTGGAGCTGCAGAGGAAACACAGCAACGCTGGAAATCCC 3571
Qy 561 GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsn 580
Db 3572 GAAATTCAGTACACTTCCAACTAACCAAGCTGTGTAATGTGGACTTTTCTGGGACACT 3631
Qy 581 AsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db 3632 AATGGCGTGTATTACAGAGCTTCGCCCATTTGGCCACCATACCTGACTGTGTAATCTG 3688
```

RESULT 8
AX282480
LOCUS

```
DEFINITION Sequence 1 from Patent WO0168888.  
ACCESSION AX282480  
VERSION AX282480.1 GI:16609611  
KEYWORDS  
SOURCE Adeno-associated virus 2  
ORGANISM Adeno-associated virus 2  
Virus; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.  
REFERENCE 1  
AUTHORS Xiao W. and Daring M.J.  
TITLE Production of chimeric capsid vectors  
JOURNAL Patent: WO 0168888-A 1 20-SEP-2001;  
Neurologix, Inc. (US)  
FEATURES  
Location/Qualifiers  
1..4679  
/organism="Adeno-associated virus 2"  
/mol_type="genomic DNA"  
/db_xref="taxon:10804" 964 t  
BASE COUNT 1198 a 1262 c 1255 g 964 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 5,44e-195 Length: 4679  
Score: 2764.50 Matches: 496  
Percent Similarity: 89.98% Conservative: 43  
Best Local Similarity: 82.80% Mismatches: 59  
Query Match: 85.04% Indels: 1  
Gaps: 1  
DB: 6  
US-09-807-802A-15 (1-599) x AX282480 (1-4679)
```

```
Qy 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSer 20  
Db 2614 ACAGCTCCGGGAAAGAGAGCGCGTAGAGCACTCTCTGTGGAGCCAGACTCTCTCTCG 2673  
Qy 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGly 40  
Db 2674 GGAACCGGAAAGCGGGCCAGCAGCTCTGCAAGAAAGATTGAATTTTGGTCAGACTCGA 2733  
Qy 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60  
Db 2734 GACGAGACTCAGTACCTGACCCCGGAGCTCTCGGACAGCCACCCAGCAGCCCTCTGT 2793  
Qy 61 ValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGly 80  
Db 2794 CTGGGAACTAATACGATGGCTACAGGCACTGGCGCACCAATGGCAGACAATAACGAGGCG 2853  
Qy 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100  
Db 2854 GCCGACGGAGTGGGTAAATTCCTCGGAAATTTGGCAATTCGATTCACATGGATGGCGAC 2913  
Qy 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyr 120  
Db 2914 AGAGTCAATCACCAACGACCCGAACTGGGCGCTCTGCGACAGCCACCTACCAACACCTCTAC 2973  
Qy 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer 140  
Db 2974 AAACAAATTTCCAGCCAATCA---GGAGCCTCGAACGACAATCACTACTTTGGCTACAGC 3030  
Qy 141 ThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160  
Db 3031 ACCCTCTGGGGGTATTTTGTACTTCACAGATTCCTACTGCCACTTTTACCAGCTGACTGG 3090  
Qy 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180  
Db 3091 CAAGACTCATCAACAACAACACTGGGGATTCGAGCCCAAGAGACTCAACTTCAAGCTCTTT 3150  
Qy 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeu 200  
Db 3151 AACATTCAAGTCAAGAGGTCAACAGATTCACGGTACGACGAGCATTTGCCAATACCTTT 3210  
Qy 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220  
Db 3211 ACCAGCAGGTTTCAAGTGTATTACTACTCGGAGTACAGCTCCCGTACGCTCTCGGCTCG 3270
```

```
Qy 221 AlahisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGly 240
Db 3271 GCGCATCAAGGATGCTTCCTCCGCGCTCCACGACAGAGCTTCATGGTGCCACATATGGA 3330

Qy 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
Db 3331 TACCTCACCTTGAACAACAGGAGTCAGGACGATAGGACGCTCTCATTTTACTGCTGGAG 3390

Qy 261 TyrPheProSerGlnMetLeuArgThrGlyVAnAsnPheThrPheSerTyrThrPheGlu 280
Db 3391 TACITTCCTCTCAGATGCTCGTACCGAAACAACTTTTACCTTCAGCTACACTTTTGAG 3450

Qy 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
Db 3451 GACGTCCTTTCACAGCAGCTAGCTCCACAGCAGAGTCTGGACCGCTCTCATGAATCCT 3510

Qy 301 LeuLeuAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
Db 3511 CTCATCGACCACTGATCTATTTACTTGACAGAACAACTCAAGTGGAAACCCACG 3570

Qy 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
Db 3571 CAGTCAAGCTTCAGTTTCTCAGCCGAGCGAGTGACATTCGGACCACTTAGGAAC 3630

Qy 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
Db 3631 TGGCTTCCTGACCCCTGTACCGCCAGCAGCAGTATCAAGACATCTCGCGATAACAAC 3690

Qy 361 AsnSerAsnPheThrTyrThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 380
Db 3691 AACAGTGAATACTCGTGACTGGAGCTACCAAGTACCACCTCAATGGCAGAGCTCTCTG 3750

Qy 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet 400
Db 3751 GTGATCCGGCCCGCCATGGCAGCCACCAAGACGATGAGAAAGTTTTTCTCTCAG 3810

Qy 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
Db 3811 AGCGGGTCTCATCTTTGGGAAGCAAGCTCAGAGAAACAAATGTGGACATGAAAG 3870

Qy 421 ValMetIleThrAspGluGluGluLysAlaThrAsnProValAlaThrGluArgPhe 440
Db 3871 GTCATGATTACAGAGAAAGAAATCAGGACAACTCCCGTGGCTACGGAGCGAT 3930

Qy 441 GlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHis 460
Db 3931 GGTTCGTATCTACCACTCCAGAGAGCAACAGACAGCTACCGCAGATGTCAAC 3990

Qy 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro 480
Db 3991 ACACAAGGCGTCTTCCAGGCATGCTCGCAGGACAGAGATGTGTACTTCAGGGGCC 4050

Qy 481 IleTrpAlaValIleProHisThrAspGlyHisPheHisProSerProLeuMetClyGly 500
Db 4051 ATCTGGGCAAGATTCACACAGCAGCAGCACTTTTCAACCTCTCTCCCTCATGGGTGGA 4110

Qy 501 PheGlyLeuLysAsnProProGlnIleLeuLysAsnThrProValProAlaAsn 520
Db 4111 TTCGGATTAAACCCCTCTCCACAGATCTCATCAAGAACACCCCGGTACCTCGGAAT 4170

Qy 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
Db 4171 CCTCGACCACTTCAGTCGGCAAGTTTGTCTCTTCATCACACAGTACTTCCACGGGA 4230

Qy 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTyrAsnPro 560
Db 4231 CAGGTCAGCGTGGAGATCAGTGGGAGCTGCAGAGGAAACAGCAACGCTGGGAATCCC 4290

Qy 561 GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsn 580
Db 4291 GAAATTCAGTACACTTCCAACTACAACAGTCTGTTAATGTGGACTTTTACTGTGGCACT 4350
```

```
581 AsnGlyLeuTyrThrGluProArgProfileGlyThrArgTyrLeuThrArgProLeu 599
Db 4351 AATGGCGTGATTATTCAGAGCCTCGCCCATTTGGCAGCAGATACCTGACTCGTAATCTG 4407

RESULT 9
AF043303 LOCUS Adeno-associated virus 2, complete genome. linear VRL 24-FEB-1998
DEFINITION AF043303
ACCESSION AF043303.1 GI:2906016
KEYWORDS Adeno-associated virus 2
SOURCE Adeno-associated virus 2
ORGANISM Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
REFERENCE 1 (bases 1 to 4679)
AUTHORS Ruffing,M., Heid,H. and Kleinschmidt,J.A.
TITLE Mutations in the carboxy terminus of adeno-associated virus 2
capsid proteins affect viral infectivity: lack of an RGD
integrin-binding motif
JOURNAL J. Gen. Virol. 75 (Pt 12), 3385-3392 (1994)
MEDLINE 95088582
PUBMED 7996133
REFERENCE 2 (bases 1 to 4679)
AUTHORS Berns,K.I., Bohenzky,R.A., Cassinotti,P., Colvin,D., Donahue,B.A.,
Dull,T., Horer,M., Kleinschmidt,J.A., Ruffing,M., Snyder,R.O.,
Tratschin,J.-D. and Weitz,M.
TITLE Direct Submission
JOURNAL Submitted (15-JAN-1998) Cell Genesys Inc., 342 Lakeside Dr., Foster
City, CA 94404, USA
FEATURES
source Location/Qualifiers
1..4679
/organism="Adeno-associated virus 2"
/mol_type="genomic DNA"
/db_xref="taxon:10804"
/notes="changes relative to the original sequence, GenBank
Accession Number J01901, have been detected and verified
by several different laboratories"
repeat_region 1..145
/notes="inverted terminal repeat"
misc_feature 42..83
/rrp_type="inverted"
precursor_RNA 287..4451
CDS join(321..1906,2228..2252)
/product="Rep 68 protein"
/protein_id="AAC03774.1"
/db_xref="GI:2906017"
/translation="MFGFYEIVIKVPSDLDEHLPGISDSFVNVAEKEWELPPDSMD
LNIEQAPLVAEKLQDFLTERVRVSKAPEALFFVQFEKGSYFHHHVLVETTVGVS
MVLGRFSQIREKLIQRIYRGIEPTLPNFAVTKTRNGAGGKNKVDECIYNYLLPK
TQELQAWNTMNEQYLSACLNTKELVAQLHTVSTQTEONKQNPNSDAPVIRS
KTSARVYELVGLVDKGTISEKOWIOBDOASYISFNAASRSQIKALDNAGKIMSL
TKTAPDYLVCQVEDISSNRIYKILELNGYDPQYASVFLGWATKKFGKRNITWLFK
PATGKTNIAEAIAHTVPFGCVNTNENPFNDKVMVWEEGQMTAKVVSAAK
ILGSKVRVDPQCKSAQIDTPVITVNTNMCADVIGNSTTTFEHOQPLQDRMPKPEL
TRLDHDFGKTKQEVKQDFERWAKDHVEHEFYVKGGAKKRPASDADISEPKV
RESVAPSTSDAEASINYADRLARGHSL"
321..2186
/codon_start=1
/product="Rep 78 protein"
/protein_id="AAC03775.1"
/db_xref="GI:2906018"
/translation="MFGFYEIVIKVPSDLDEHLPGISDSFVNVAEKEWELPPDSMD
LNIEQAPLVAEKLQDFLTERVRVSKAPEALFFVQFEKGSYFHHHVLVETTVGVS
MVLGRFSQIREKLIQRIYRGIEPTLPNFAVTKTRNGAGGKNKVDECIYNYLLPK
TQELQAWNTMNEQYLSACLNTKELVAQLHTVSTQTEONKQNPNSDAPVIRS
KTSARVYELVGLVDKGTISEKOWIOBDOASYISFNAASRSQIKALDNAGKIMSL
TKTAPDYLVCQVEDISSNRIYKILELNGYDPQYASVFLGWATKKFGKRNITWLFK
PATGKTNIAEAIAHTVPFGCVNTNENPFNDKVMVWEEGQMTAKVVSAAK
ILGSKVRVDPQCKSAQIDTPVITVNTNMCADVIGNSTTTFEHOQPLQDRMPKPEL
TRLDHDFGKTKQEVKQDFERWAKDHVEHEFYVKGGAKKRPASDADISEPKV
RESVAPSTSDAEASINYADRLARGHSL"
321..2186
/codon_start=1
```


Db 2614 ACGGCTCCGGGAAAAAGAGCGGTAGAGCACTCTCTGTGGAGCCAGACTCTCTCTCG 2673
Qy 21 GlyIleGlyThrGlyGlnGlnProAlaLysAlaGlyLeuAspPheGlyGlnThrGly 40
Db 2674 GGAACGGGAAGCGGCGCCGAGCGCTCGAGAAAGATTGATTTGGTCAGACTGA 2733
Qy 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaLa 60
Db 2734 GACGCAGACTCAGTACTGACCCCGACGCTCTCGGACAGCCACAGCAGCCCTCTGT 2793
Qy 61 ValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnGluGly 80
Db 2794 CTGGGAACCTAATACGATGCTACAGCAGTGGCGCACCAATGGCAGCAATAACAGGGC 2853
Qy 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
Db 2854 GCCGACGGAGTGGTAAATCTCTCGGGAATTTGGCATTGGATTCCACATGGATGGCGAC 2913
Qy 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTrpAsnAsnHisLeuTyr 120
Db 2914 AGAGTCATCACCACAGCACCGGACCTGGGCGCTGCCACCTCAACAACACCACTCTAC 2973
Qy 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer 140
Db 2974 AACAAATTTCCAGGCCAATCA--GGAGCCTCGAACGACATCACTACTTTGGCTACAGC 3030
Qy 141 ThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
Db 3031 ACCCTTGGGGTATTTTGCATTCACAGATTTCCACTGCCACTTTTCCACAGTCACTGG 3090
Qy 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
Db 3091 CAAGACTCATCAACAACTGGGGATTCGACCCCAAGAGACTCAACTTCAAGTCTCTT 3150
Qy 181 AsnIleGlnValLysGluValThrAsnAspGlyValThrIleAlaAsnAsnLeu 200
Db 3151 AACATTCAGTCAAGAGGTTCGCGAATGACGATGACGACAGATTCCTCAATACCTT 3210
Qy 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
Db 3211 ACCAGCAGGTTTCAGGTGTTACTGACTCGGAGTACCAGTCCCGTACGCTCTCGGCTCG 3270
Qy 221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGly 240
Db 3271 GCGCATCAAGGATGTCCTCCGCGCTTCCAGCAGACGCTCTCATGGTGCACAGTATGA 3330
Qy 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
Db 3331 TACTTCACCTGAAACACGGAGTCAGGACGATAGGACGCTCTTCATTTTACTGCTGGAG 3390
Qy 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
Db 3391 TACTTTCTCTCAGATGCTGGTACCGGAAACAACTTTACCTTCAGTACACTTTTGAG 3450
Qy 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
Db 3451 GACGTTCTTTCACAGCAGCTACGCTCAGCAGAGTCTGGACCGCTCTCATGAATCT 3510
Qy 301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
Db 3511 CTCTCGACAGTACTGTTATCTTCTGAGCAGACAAACACTCCAGAGTGAACACCAAGC 3570
Qy 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
Db 3571 CAGTCAAGGCTTCAGTTTCTCAGGCGCGGAGGTGACATTCGGGACCACTAGGAAC 3630
Qy 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
Db 3631 TGCGTCTCTGGACCTTGTACCGCCAGCAGCGAGTATCAAGACATCTCCGGATACCAAC 3690
Qy 361 AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 380
Db 3691 AACAGTGAATACTCTGTTGGAGTGGAGCTACCAAGTACCACTCAATGGCAGACTCTCTG 3750

Qy 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet 400
Db 3751 GTGAATCCGGCCCGCGCATGTGCAAGCACCAAGGACGATGAAGAAAAGTTTTTCTCTCAG 3810
Qy 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
Db 3811 AGCGGGGTTCATCTCTTTGGGAAGCAAGGCTCAGAGAAAACAAATGTGGACATTTGAAAG 3870
Qy 421 ValMetIleThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
Db 3871 GTCATGATTACAGACAGAGAGAAATCAGACACCAATCCCGTGGCTTACCGAGCAGTAT 3930
Qy 441 GlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHis 460
Db 3931 GGTTCTGTATCTCAACACTCCAGAGAGCAACAGACAGCAGCTACCGCAGATGTCAAC 3990
Qy 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro 480
Db 3991 ACACAAGCGCTTCTTCCAGGCGATGCTCGGACAGACAGATGTGTACTTTCAGGGGCCC 4050
Qy 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
Db 4051 ATCTGGGCAAGATTTCCACACCGACGAGCATTTTCACTCTCTCCCTCATGGGTGA 4110
Qy 501 PheGlyLeuLysAsnProProProGlnIleLeuIleLysAsnThrProValProAlaAsn 520
Db 4111 TTCGACTTAAACACCTCTCCACAGATTTCTCATCAAGAACACCCCGGTACCTGCGAAT 4170
Qy 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
Db 4171 CTTTCGACCACTTCAGTGGCGCAAGTTTCTCTTCTCATCACACAGTATCTCCACGGGA 4230
Qy 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPro 560
Db 4231 CAGTTCAGCTGGAGATGAGTGGAGCTGAGAGGAAACAGACAAACGCTGGATCCC 4290
Qy 561 GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAsnValAspPheThrValAspAsn 580
Db 4291 GAAATTCAGTACACTTCCACTACACAAAGTCTGTAAATGTGGACTTTTACTGTGGACACT 4350
Qy 581 AsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db 4351 AATGGCGTATTATCAGAGCCTCGCCCCATTGGCACCAGATACCTGACTGTAATCTG 4407
RESULT 10
LOCUS AF369963 7327 bp DNA circular SYN 22-JUL-2001
DEFINITION Cloning vector pAAV-RC, complete sequence.
ACCESSION AF369963
VERSION AF369963.1 GI:14994089
KEYWORDS
SOURCE
ORGANISM
Cloning vector pAAV-RC
Cloning vector pAAV-RC
artificial sequences; vectors.
REFERENCE
1 (bases 1 to 7327)
PAAV-RC for AAV Helper-Free System
Grafsky, A.J. III.
TITLE
JOURNAL
AUTHORS
REFERENCE
2 (bases 1 to 7327)
Grafsky, A.J. III.
Direct Submission
TITLE
JOURNAL
AUTHORS
SUBMITTED (16-APR-2001) Technical Services, Strategene, 11011 N.
Torrey Pines Rd., La Jolla, CA 92037, USA
FEATURES
Location/Qualifiers
1..7327
/organism="Cloning vector pAAV-RC"
/mol_type="genomic DNA"
/db_xref="taxon:161364"
/lab_host="Escherichia coli"
/note="used for recombinant DNA processing"
131..1996
/note="AAV-2 rep"
CDS

```
/codon start=1
/transl table=1
/product="adeno-associated virus 2 rep protein"
/protein_id="AAK76418.1"
/db_xref="GI:14994090"
/translation="MPGFYEIVIKVPSDLDELHPGLISDSFVNWVAEKELWPPDDSDMD
LNLIEQAPLTVAKLQRLDFLWRRVSKAPEALFFVQPEKGSYFHMVLYVETTCVKS
MVLGRPLSQIRKQLQRIYRGIEPTLPWFATKTRNGAGGKGVKVDYIYNYLLPK
TPELQWANNMEQILASCLNTERKLVQAHLTHVSTQEQKNGKPNPSDAPVRS
KTSARYMELWLDVKGITSEKOWIQEDQASYISFNAASNSRSQKALDNNAGKIMSL
TKTAPYLVGQOPVEDISSNRIYKILELNGYDQYAAAFVFLGMAFKFKRNTIWLFG
PATTKNTIAEAIATHTVPYGCNVNENPFNDCVDMKVIWMBEGKMTAKVBSAKA
ILGSKVRVDOKSSAQIDPPTVPIVTNTNCNCAVIDGNSITTFHQOPLQDRMEFEL
RLRHDHFGKTVKQEVKDFRWAKOHVVEHEFYVKKGAKKRPAPSDADISEKRV
RSVQAQPSDSEASINADRYQNKCSRHVGNLMLFPQRCERNQNSNICTFGQK
DCECFPVSESPQSVVKKAYQKLCYIIHMGKVPDACTACDCLVNVDLDDCIFEQ"
2013..4220
/notes="cap VP1"
/codon start=1
/transl table=1
/product="capsid protein VP1"
/protein_id="AAK76419.1"
/db_xref="GI:14994091"
/translation="MAADGYLPDWLEDTLSEGIROWMKLPGPPPPKPAERHKDSRG
LVLPGYKLPNGLDKGEVNEADAALAEHDKAYDRLDSDGNPVLKYNHADAEFOE
RLKEDTSPGNLGRAVFOAKGVLEPLGLVESPVTAPGKRPVHSVPEDSSSGTG
KAGQOPARKLAFPGQTGADSDVPDQPLQPPAAPSGGLGTNTMATGSGAPMADNNEGA
DGVNSGNHWCSTWMDRVIITSTWALPTNNHLYKQISSQSGASNDNHYFGYS
TPWGYDFNRHFCHFSRDMQRLNNWGFPRKFLNFKLVNQVEVTQNDGTTIAN
NLTSVQVFTDSEYQLPYVLGAHQCLPFPADVPMVQYGLYLNNGSQAVGRSSF
YCLEYPPSMLRTGNNTFSYTFEDVPHSSVAHSOSLDRLNPLIDYLYLSTRNT
PSTTQWRLQFSOAGSDIRDOGRNMLPGCVROORVSKTSADNNNSYSVWTGATKY
HLNGRSLVNPGMASHKDDKEKFPFQSGVLLFGQSEKTNVDIEKMTNDESEIR
TNFVATEQTVGSVSTNLRGNQAADVNTQVLPGMVWQDRDVLQGLVWAKLPH
DGHFPLPLMGFLKPPQILKNTVPANPSTTFAAKFASFTQYSTQGVSEI
EWELQKENSRRWPEIQYTSNYSNVNVDFTVDNNGVSEPRPIGTLYLRLN"
4705..5161
/notes="fl origin"
5292..6152
/notes="confers ampicillin resistance"
/codon start=1
/transl table=1
/product="beta-lactamase"
/protein_id="AAK76420.1"
/db_xref="GI:14994092"
/translations="MSIQHFRVALIFPFAAFCLPVPFAHPETLVKVKDAEDQLGARVGY
IEDLNSGKILSEPRPEERFPMMSTFKLLCGAVLSRIIDAGQELGRRIHYSDNDLVE
YSPVTEKHLDGTMTVRELCSAAITMSDNTAANLLLTIGGPKELTAFLNMGDHYVRL
DRWEPELNAIPNDERDITTPVAMATTLRKALTGELLTLASRQQLIDMWEADKVAGPL
LRGALPAGWFIADKSGAGERSGIIAALGPLDGPGRSIVIVITGTSQATWMDERNRQIA
EIGASLIKHW"
6161..7019
/notes="ColE1 origin"
BASE COUNT 1903 a 1887 c 1912 g 1625 t
ORIGIN
```

```
Alignment Scores:
Pred. No.: 9,95e-195 Length: 7327
Score: 2764.50 Matches: 496
Percent Similarity: 89.98% Conservative: 43
Best Local Similarity: 82.80% Mismatches: 59
Query Match: 85.04% Indels: 1
DB: 12 Gaps: 1
```

US-09-807-802a-15 (1-599) x AF369963 (1-7327)

```
Qy 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSer 20
Db 2424 ACGGCTCCGGGAAAAAGAGCCGGTAGAGCACTCTCTCTGTGGAGCAGACTCTCTCTCG 2483
Qy 21 GlyIleGlyThrGlyGlnGlnProAlaLysLysArgLeuAsnPhelGlnThrGly 40
Db 2484 GGAACCGGAAAGCGGCCAGACGCTGCAAGAAAAGATTTGAATTTTGGTCAGACTGGA 2543
```

```
Qy 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
Db 2544 GACGCACTCAGTACCTGAGCCCCAGCCTCTCGAGCAGCCACCCAGCAGCCCCCTCTCGT 2603
Qy 61 ValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGly 80
Db 2604 CTGGGAACATAATACGATGGCTACAGGAGTCGCGCACCACCAATGGCAGACAATAACGAGGC 2663
Qy 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
Db 2664 GCGGACGAGTGGGTAAATCTCTCGGAAATGGCAATTCGCAATTCGCAATGGATGGCGCAC 2723
Qy 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTrpAsnAsnHisLeuTrp 120
Db 2724 AGAGTCATCACACACGACCCGAAACCTTGGGCCCTTGGCCACCTACACACACCACTCTTAC 2783
Qy 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTrpPheGlyTrpSer 140
Db 2784 AAACAAATTTCCAGCCCAATCA--GGAGCCTCGAACGACAACTCACTACTTTGGCTACAGC 2840
Qy 141 ThrProTrpGlyTrpPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
Db 2841 ACCCTCTGGGGGTATTTTGACTTCAACAGATTCCACTGCCACTTTTTCACCACTGACTGG 2900
Qy 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPhelLysLeuPhe 180
Db 2901 CAAGACTCATCAACACACTGGGGATTCGACCCCAAGAGACTCACTTTCAAGCTCTTT 2960
Qy 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeu 200
Db 2961 AACATTCAAGTCAAAAGAGTCCGCGAGAAATGACGGTACGACGAGCATTTGCCAATAACCTT 3020
Qy 201 ThrSerThrValGlnValPheSerAspSerGluTrpGlnLeuProTrpValLeuGlySer 220
Db 3021 ACCAGCAGGTTTCAAGTGTCTTACTGACTCGGAGTACCACTCCCGTACGCTCTCGGCTCG 3080
Qy 221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTrpGly 240
Db 3081 GCGCATCAAGAGATGCTCCGCGCTTCCAGCAGACAGCTTTCATGCTGGTCCACAGTATGGA 3140
Qy 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTrpCysLeuGlu 260
Db 3141 TACCTTCACTTCAACCAAGAGTCCAGGAGTACGAGTACGCTTCTTCTTCTTCTGCTGGAG 3200
Qy 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTrpThrPheGlu 280
Db 3201 TACTTTCTTCTCAGATCTCGGTACCGGAAACAACTTTTACCTTCACTTCACTTCTTCTGAG 3260
Qy 281 GluValProPheHisSerSerTrpAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
Db 3261 GAGCTTCTTCTTCCAGCAGCAGTACGCTCACAGCCAGAGTCTGGACCGTCTCATGATCTCT 3320
Qy 301 LeuIleAspGlnTrpLeuTrpTrpLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
Db 3321 CTCTCGACCACTACCTGTATTACTTGTAGCAGAACAAACACTCCCAAGTGGAAACCAACG 3380
Qy 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
Db 3381 CAGTCAAGGCTTCACTTTTCTCAGGCGCGGAGGAGTACATTCGGGACCACTGCTAGGAAC 3440
Qy 341 TrpLeuProGlyProCysTrpArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
Db 3441 TGGCTTCTTGGACCTGTATACCGCCAGCAGGAGTATCAAGAGCATCTGCGGATAACAC 3500
Qy 361 AsnSerAsnPheThrTrpThrGlyAlaSerLysTrpAsnLeuAsnGlyArgGluSerIle 380
Db 3501 AACAGTGAATACTCTGGAGTGGAGTACCAAGTACCACTCAATGGCAGAGACTCTCTG 3560
Qy 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet 400
Db 3561 GTGAATTCGGGCGCCCGCCATGCGCAAGCCCAAGGAGCATGAAGAAAGATTTTTTCTCTCAG 3620
```

```
Qy 401 SerGlyValMetIlePheGlySerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
Db 3621 AGCGGGTTCATCTTTGGAGCAAGGCTCAGAGAAAACAATGTGGACATTTGAAAG 3680
Qy 421 ValMetIleThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
Db 3681 GTCATGATTACAGACGAGAGGAAATCAGGACAACTATCCCGTGGCTACGGACAGTAT 3740
Qy 441 GlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHis 460
Db 3741 GGTTCTGTATCTACCAACCTCCAGAGAGCAACAGACAAAGCAGCTACCGCAGATGTCAAC 3800
Qy 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro 480
Db 3801 ACACAGCGCTTCTTCAGAGCATGTCTGGCAGACAGAGATGTACTCTTCAGGGGCC 3860
Qy 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
Db 3861 ATCTGGGCAAGATTCACACAGCAGGACATTTTCACCCCTCTCCCTCATGGGTGA 3920
Qy 501 PheGlyLeuLysAsnProProGlnIleLeuLysAsnThrProValProAlaAsn 520
Db 3921 TTCGAGCTTAACACCCCTCTCCACAGATTTCTCATCAAGAACACCCCGGTACCTGCGAAT 3980
Qy 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
Db 3981 CCTTCGACCACTTCAGTGGCGGCAAGTTGCTTCCTTCATCACACAGTACTCCACGGGA 4040
Qy 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPro 560
Db 4041 CAGTTCAGCGTGGAGATCGAGTGGGAGCTGCAGAGGAGAAACAGCAACGCTGGAAATCCC 4100
Qy 561 GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsn 580
Db 4101 GAAATTCAGTACACTTCCAACTACAAAGTCTGTAAATGTGGACTTTTACTGTGGACACT 4160
Qy 581 AsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db 4161 AATGCGGTGATTCAGAGCCTCGCCCATTTGGCACCAGATACCTGACTCGTAATCTG 4217

RESULT 11
LOCUS AR222045 7557 bp DNA linear PAT 26-SEP-2002
DEFINITION Sequence 3 from patent US 6429001.
ACCESSION AR222045
VERSION AR222045.1 GI:23329401
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7557)
AUTHORS Hardy, S.F.
TITLE Recombinant AAV packaging systems
JOURNAL Patent: US 6429001-A 3 06-AUG-2002;
FEATURES Location/Qualifiers
source 1..7557
/organism="unknown"
BASE COUNT 1970 a 1940 c 1953 g 1694 t
ORIGIN

Alignment Scores:
Pred. No.: 1,04e-194 Length: 7557
Score: 2764.50 Matches: 496
Percent Similarity: 89.98% Conservative: 43
Best Local Similarity: 82.80% Mismatches: 59
Query Match: 85.04% Indels: 1
DB: 6 Gaps: 1

US-09-807-802A-15 (1-599) x AR222045 (1-7557)

Qy 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSer 20
Db 2642 ACGGCTCCGGGAAAAAGAGCGCGGTAGAGCACTCTCTCTGTGGACGACGACTCTCTCTG 2701
```

```
Qy 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGly 40
Db 2702 GGAACCGGAAAGCGGCGCCAGAGCCTGCAAGAAAAAGATTGAATTTTGGTTCACAGCTGGA 2761
Qy 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
Db 2762 GAGCAGACTCAGTACTGACCCCGAGCCTCTCGACAGCACCAGCAGCCCCCTCTGGT 2821
Qy 61 ValGlyProThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGlyGly 80
Db 2822 CTGGAACTAATACGATGGCTACAGGCAGTGGGCGCACCAATGGCAGACAATAACAGAGGC 2881
Qy 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
Db 2882 GCCGACGGAGTGGGTAAATCTCTCGGAAATTTGGCATTTCCGATTTCCACATGGATGGGCGAC 2941
Qy 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyr 120
Db 2942 AGAGTCATCACACCAGCAGCCCGAACCTGGCGCCCTGACCACTACAAACACCACCTCTAC 3001
Qy 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer 140
Db 3002 AAACAAATTTCCAGCCCAATCA--GGAGCCTCGAACGACAATCACTACTTTGGCTCAGC 3058
Qy 141 ThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
Db 3059 ACCCTTGGGGGTATTTTGACTTCAACAGATTCACCTGCCACTTTTCCACACGCTGACTGG 3118
Qy 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
Db 3119 CAAAGACTCATCAACAACAATCGGGATTCCGACCCCAAGAGACTCAACTTCAAGCTCTTT 3178
Qy 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeu 200
Db 3179 AACATTCAGTCAAGAGGTCAAGCAGATACGCGTACGACGACGATGGCCAAATAACCTT 3238
Qy 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
Db 3239 ACCGACGGTTTCAAGTGTATTTACTGACTCGAGTACCGAGTCCCGCTCCGTACTCTGGCTCG 3298
Qy 221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGly 240
Db 3299 GCGCATCAAGATGCTCCCGCGTCCAGCAGACAGCTTCTCATGGTGCCACAGTATGGA 3358
Qy 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
Db 3359 TACCTCACCTCGAAACAACGGGAGTCAGGACAGTACGAGCGCTCTCATTAATCTGGCTGGAG 3418
Qy 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
Db 3419 TACTTTCTTCTCAGATGCTGCGTACCGGAAACAACCTTTTACCTTCAGCTACACTTTTTCAG 3478
Qy 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
Db 3479 GACGTCTCTTCCACAGCAGCTACGCTCACAGCCAGAGTCTGGACCGCTCTCATGAATCCT 3538
Qy 301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
Db 3539 CTCATCGACCATGATACCTGTATTACTTGAGCAGAAACAACACTCCCAAGTGGAAACCCACG 3598
Qy 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
Db 3599 CAGTCAAGGCTTCAGTTTCTCAGCGCGGAGCGAGTGACATTCGGGACCATGCTAGGAAC 3658
Qy 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
Db 3659 TGGCTTCTGACCTGTATTACCGCAGCAGCGAGTATCAAAGACATCTCGCGAATAACAAC 3718
Qy 361 AsnSerAsnPheThrThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 380
Db 3719 AACAGTGAATACTCTGGACTGGAGCTACCAGTACCACCTCAATGGCAGAGACTCTCTG 3778
```

```

QY 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspGluAspLysPhePheProMet 400
DB 3779 GTGAATCCGGGCGCCGCAATGGCAAGCAAGAGCGATGAAGAAAGTTTTTCTCAG 3838
QY 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
DB 3839 AGCGGGGTTCTCATCTTTGGGAAGCAAGCTCAGAGAAACAAATGTGGACATTGAAG 3898
QY 421 ValMetIleThrAspGluGluLysAlaThrAsnProValAlaThrGluArgPhe 440
DB 3899 GTCATGATTACAGCAAGAGGAAATCAGACAAACCAATCCCGTGGCTACGGAGCAGTAT 3958
QY 441 GlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHis 460
DB 3959 GGTTCGTATCTACCAACCTCCAGAGGCAACAGACAGAGCTACCCAGATGTCAAC 4018
QY 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTrpLeuGlnGlyPro 480
DB 4019 ACACAAGGGGTTCTCCAGGCATGGTCTGGCAGGACAGAGATGTGTACCTTCAGGGGCC 4078
QY 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
DB 4079 ATCTGGGCAAGATTCACACACGACGACGACATTTTCAACCCCTCTCCCTCATGGGTGA 4138
QY 501 PheGlyLeuLysAsnProProGlnIleLeuLysAsnThrProValProAlaAsn 520
DB 4139 TTGGGACTTAAACCCCTCTCCAGAGTTCTCATCAAGAACACCCCGGTACTCTGGCAAT 4198
QY 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTrpSerThrGly 540
DB 4199 CCTTCGACCACTTCAGTGGCGGCAAGTTTGTCTTCTCATCACAGTACTCCACGGGA 4258
QY 541 GlnValSerValGluLysGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPro 560
DB 4259 CAGGTGAGGTGAGATCGAGTGGGAGCTGCAGAGGAAACAGCAACGCTGGAATCCC 4318
QY 561 GluValGlnTrpThrSerAsnTrpAlaLysSerAlaAsnValAspPheThrValAspAsn 580
DB 4319 GAATTCAGTACACTTCCAACTACACAGTCTGTTAAATGTGGACTTACTGTGGACACT 4378
QY 581 AsnGlyLeuTrpThrGluProArgProIleGlyThrArgTrpLeuThrArgProLeu 599
DB 4379 AATGGGCTGATTTCAGAGCTCGCCCATTTGGCACCAGATACCTGACTCGTAATCTG 4435

RESULT 12
AX205073 LOCUS 7557 bp DNA linear PAT 30-AUG-2001
DEFINITION Sequence 3 from Patent WO015361.
ACCESSION AX205073
VERSION AX205073.1 GI:15394314
KEYWORDS unidentified
SOURCE unidentified
ORGANISM unclassified.
REFERENCE
1 Hardy, S.F.
AUTHORS Recombinant aav packaging systems
TITLE Patent: WO 015361-A.3 02-AUG-2001;
JOURNAL Chiron Corporation (US)
FEATURES
Location/Qualifiers
source
1. 7557
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/notes="recombinant DNA"
BASE COUNT 1970 a 1940 c 1953 g 1694 t
ORIGIN

Alignment Scores:
Pred. No.: 1,04e-194 Length: 7557
Score: 2764.50 Matches: 496
Percent Similarity: 89.98% Conservative: 43
Best Local Similarity: 82.80% Mismatches: 59

```

```

Query Match: 85.04% Indels: 1
DB: 6 Gaps: 1
US-09-807-802A-15 (1-599) x AX205073 (1-7557)
QY 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSer 20
DB 2642 ACGGCTCCGGGAAAAAGAGGCGCGGTAGAGCACTCTCTGTGGAGCCAGACTCCTCTCG 2701
QY 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGly 40
DB 2702 GGAACCGGAAAGCGCGGCAGCAGCTTCAAGAAAAAGATTGAATTTTGGTCAGACTGA 2761
QY 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
DB 2762 GAGCAGACTCTAGTACCTAGCCCCAGCCTCTCGAGACGCCACAGCAGCCCCCTCTGTG 2821
QY 61 ValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGly 80
DB 2822 CTGGGAACCTAATACGATGGCTACAGCGCAGTGGCGCAACCAATGGCAGACAATAACGAGGC 2881
QY 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
DB 2882 GCGGACGAGTGGGTAAATTCCTCGGGAATTTGGCATTTGGCATTCACATGGATGGCGCAG 2941
QY 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyr 120
DB 2942 AGAGTCATCACACACGACCGACCCGAACCTGGGCCCTGCCACCTACACACACCTCTAC 3001
QY 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer 140
DB 3002 AAACAAATTTTCAGCCAATCA---GGAGCCTCGAACGACCAATCACTACTTTGGCTACAGC 3058
QY 141 ThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
DB 3059 ACCCTCTGGGGTATTTTGGCTTCAACAGATTCACATGGCCACTTTTCCACGACTGACTGG 3118
QY 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
DB 3119 CAAGACTCATCAACAACACTGGGGATTCGACCCCAAGAGACTCAACTTCAAGCTCTTT 3178
QY 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeu 200
DB 3179 AACATTCAGTCAAGAGGTACGCGAGATGACGGGTACGACGACGATTCGCCAATACCTT 3238
QY 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
DB 3239 ACCGACGCGTTTCAGGTGTTTACTGACTCGGAGTACCAGCTCCCGTACGCTCTCTCGGCTCG 3298
QY 221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGly 240
DB 3299 GCGCATCAAGGATGCTCCCGCGTTCCCGCAGCAGCGCTTTCATGTGTGCGCACAGTATGA 3358
QY 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValAlaGlyArgSerSerPheTyrCysLeuGlu 260
DB 3359 TACCTCACCCCTGAACAACCGGAGTCAGGCGAGTACGACGCTTCTTCAATTTTACTGCTGGAG 3418
QY 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnAsnPheThrPheSerTyrThrPheGlu 280
DB 3419 TACTTTCTTCTCAGATGCTGCTACCGGAAACAACTTTTACCTTCAGCTACACTTTTGGAG 3478
QY 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
DB 3479 GAGGTTTCTTCCACAGCAGCTACGCTCACAGCCAGAGTCTGGACCGCTCTCATGAATCCT 3538
QY 301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
DB 3539 CTCATCGACCTACCTGTATTACTTTGAGCAGAACAAACACTCCCAAGTGTGAACACGACG 3598
QY 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
DB 3599 CAGTCAAGGCTTCAGTTTTTCTCAGGCCGAGGAGTGACATTCGGGACCGACTAGGNAC 3658

```

```
Qy 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
Db 3659 TGGCTTCTGGACCTGTTACCGCCAGCAGCGAGTATCAAGACATCTGCGGATAACAAC 3718
Qy 361 AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 380
Db 3719 AACAGTGAATACTCGTGAGCTGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTG 3778
Qy 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet 400
Db 3779 GTGAATCCGGCCCGCCCATGCGCAAGCCACAGGACGATGAAGAAAGTTTTCCTCAG 3838
Qy 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
Db 3839 AGCGGGTCTCATCTTTGGGAAGCAGGCTCAGAGAAAACAATGTGGACATTTGAAAG 3898
Qy 421 ValMetIleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
Db 3899 GTCATGATTACAGACGAAAGGAAATCAGGAAACAACCAATCCCGTGGCTACCGGACGAT 3958
Qy 441 GlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHis 460
Db 3959 GGTTCGTGATCTACCACTCTCAGAGAGCAACAGACAGCAGTACCGCAGATGTCAAC 4018
Qy 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro 480
Db 4019 ACACAAGCGGTTCTTCCAGGCATGCTCTGGCAGGACAGAGATGTGTACTCTTCAGGGGCC 4078
Qy 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
Db 4079 ATCTGGCAAGATTTCCACACGAGCGGACATTTTCCACCCCTCTCCCTCATGGGTGA 4138
Qy 501 PheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsn 520
Db 4139 TTCGACTTAACACCTCTCCACAGATTTCTCATCAAGACACCCCGGTACTTGGGAT 4198
Qy 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
Db 4199 CCTTCGACCCCTTCAGTGGCGCAAGTTTGTCTCTTCATCACACAGTACTTCCACGGGA 4258
Qy 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPro 560
Db 4259 CAGGTGCGGTGAGATCGATGGGAGCTGCGAAGAGAAACAGCAACGCTGGAATCCC 4318
Qy 561 GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsn 580
Db 4319 GAAATTCAGTACTCCACTCAACAGTCTGTTATGTGGACTTTACTGTGGACACT 4378
Qy 581 AsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db 4379 AATGGGTGATTTACAGAGCTCGCCCATTTGGCACCAGATACCTGACTCGTAATCTG 4435

RESULT 13
AR222044
LOCUS AR222044 8698 bp DNA linear PAT 26-SEP-2002
DEFINITION Sequence 2 from patent US 6429001.
ACCESSION AR222044
VERSION AR222044.1 GI:23329400
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS Hardy, S. F.
TITLE Recombinant AAV packaging systems
JOURNAL Patent: US 6429001-A 2 06-AUG-2002;
FEATURES Location/Qualifiers
source 1..8698
/organism="unknown"
BASE COUNT 2117 a 2368 c 2306 g 1907 t
ORIGIN
Alignment Scores:
```

```
Pred. No.: 1,25e-194 Length: 8698
Score: 2764.50 Matches: 496
Percent Similarity: 89.98% Conservative: 43
Best Local Similarity: 82.80% Mismatches: 59
Query Match: 85.04% Indels: 1
DB: 6 Gaps: 1

US-09-807-802A-15 (1-599) x AR222044 (1-8698)

Qy 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSer 20
Db 2614 ACGGCTCCGGGAAGAAAGAGCCGGTAGACACTCTCTCTGGAGCCAGACTCTCTCTCG 2673
Qy 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGly 40
Db 2674 GGAACCGGAAGCGCGCCAGCAGCTCTCAAGAAAAGATTGAATTTTGGTTCAGACTGGA 2733
Qy 41 AsnSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
Db 2734 GAGCAGACTCAGTACCTGACCCCGCAGCTCTCGGACAGCCACGAGCCCTCTCTGGT 2793
Qy 61 ValGlyProThrThrMetAlaSerGlyGlyValAlaProMetAlaAspAsnAsnGluGly 80
Db 2794 CTGGAACTAATACGATGGCTCAGGCACTGGCGCACCAATGGCAGACCAATTAACGAGGC 2853
Qy 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
Db 2854 GCCGACGAGTGGGTAACTCTCGGAAAATGGCATTTGCATTCACATGATGGCGGAC 2913
Qy 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyr 120
Db 2914 AGAGTCTATCACCAACAGCACCCTGGGCGCTGCGCCACCTTACAAACACCACTCTAC 2973
Qy 121 LysGlnIleSerSerAlaSerThrGlyValaSerAsnAspAsnHisTyrPheGlyTyrSer 140
Db 2974 AAACAATTTCCAGCCCAATCA--GGAGCTCGAACGACCAATCACTACTTTGGCTACAGC 3030
Qy 141 ThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
Db 3031 ACCCTTGGGGTATTATTGACTTCAACAGATTCACATTCGACCTCCCACTTCCACGAGTGG 3090
Qy 161 GlnArgLeuIleAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
Db 3091 CAAAGACTCATCAACAACACTGGGATTCGAGCCCAAGAGACTCAACTTCAAGCTCTTT 3150
Qy 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeu 200
Db 3151 AACATTCAGTCAAAAGAGGTCAAGCAGAAATGACGGTACGACGACGATTGCCAATAACCTT 3210
Qy 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
Db 3211 ACCAGCACGGTTCAAGTGTATTACTGACTCGAGGTACCAGCTCCCGTACGTCTCTCGGCTG 3270
Qy 221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGly 240
Db 3271 GCGCATCAAGGATGCTCCCGCGGTTCCTCCAGCAGACGCTTTCATGGTGCCACAGTATGGA 3330
Qy 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
Db 3331 TACCTCACTGAAACAACGGGAGTCAGCAGTAGGACGCTCTTCATTTACTGCTGGAG 3390
Qy 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
Db 3391 TACTTTCTCTCAGATGCTGCGTACCGGAAACAACCTTTACCTTTCAGCTACACTTTTGAG 3450
Qy 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuLeuAspArgLeuMetAsnPro 300
Db 3451 GACGTTCTTTCCACAGCAGCTACGCTCAGACGCGAGTCTGGACCGCTCTCATGAATCCT 3510
Qy 301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
Db 3511 CTGATCGACCGATACCTGTATTACTTGGGCGAACAACACTCCCAAGTGGGAACCCACCG 3570
```

```
QY 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
Db 3571 CAGTCAAGGCTTCAGTTTCTCAGGCGGAGGAGTGCATTCGGGACGACGTCTAGGAAC 3630
QY 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
Db 3631 TGGCTTCCTGGACCTGTTTACCGCCAGCAGCGAGTATCAAGACATCTCGCGATTAACAAC 3690
QY 361 AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 380
Db 3691 AACAGTGAATACCTCGTGGAGCTGAGCTACCAAGTACCACTCAATGGCAGACAGCTCTG 3750
QY 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspGluAspLysPhePheProMet 400
Db 3751 GTGAATCCGGGCGCGCCATGGCAGCCACAGGAGCATGAGAAAAGTTTTTCTCTCAG 3810
QY 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
Db 3811 AGCGGGGTTCTCATCTTTGGGAAGCAAGGCTCAGAGAAAACAAATGTGGACATTTGAAAAG 3870
QY 421 ValMetIleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
Db 3871 GTCATGATTACAGCAAGAGGAAATCAGGACAAACCAATCCCGTGCTACGGAGCAGTAT 3930
QY 441 GlyThrValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHis 460
Db 3931 GGTTCGTATCTACCAACCTCCAGAGAGGCAACAGACAGCAGTACCCGAGATGTCAAC 3990
QY 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTrpLeuGlnGlyPro 480
Db 3991 ACACAAAGGCGTTCTTCAGGCGATGCTGCGAGGACAGAGATGTGTACCTTCAGGGGCC 4050
QY 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
Db 4051 ATCTGGGCAAGAAATTCACACAGCGAGGACATTTTCACCCCTCTCCCTCTCATGGGTGA 4110
QY 501 PheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsn 520
Db 4111 TTCCGACTTAAACACCTCTCCACAGATTCCTCATCAAGAACACACCCCGGTACTCGGAAT 4170
QY 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
Db 4171 CCTTCGACCACTTCAGTGGCGAAAGTTTGTCTCTCTCATCACAGTACTCCACGGGA 4230
QY 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPro 560
Db 4231 CAGTCAGCGTGAGATCGAGTGGAGCTGCAGAGGAAACAGCAACGCTGGAATCCC 4290
QY 561 GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsn 580
Db 4291 GAAATTCAGTACACTTCCAACTACAAGTCTGTTAATGTGGACTTTACTGTGGACACT 4350
QY 581 AsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db 4351 AATGGCGTGTAATTCAGAGCCCTCCCGCATTTGGCAGCAGATACCTGATCGTAATCTG 4407
```

RESULT 14

```
AX205072
LOCUS AX205072 8698 bp DNA linear PAT 30-AUG-2001
DEFINITION Sequence 2 from Patent WO0155361.
ACCESSION AX205072
VERSION AX205072.1 GI:15394313
KEYWORDS
SOURCE unidentified
ORGANISM unidentified
REFERENCE
1 Hardy, S.F.
AUTHORS Recombinant aav packaging systems
TITLE Patent: WO 015361-A 2 02-AUG-2001;
JOURNAL Chiron Corporation (US)
FEATURES
source 1..8698
```

```
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/notes="recombinant DNA"
BASE COUNT 2117 a 2368 c 2306 g 1907 t
ORIGIN
Alignment Scores:
Pred. No.: 1,25e-194 Length: 8698
Score: 2764.50 Matches: 496
Percent Similarity: 89.98% Conservative: 43
Best Local Similarity: 82.80% Mismatches: 59
Query Match: 85.04% Indels: 1
Gaps: 6
US-09-807-802A-15 (1-599) x AX205072 (1-8698)
```

```
QY 1 ThrAlaProGlyLysLysLysArgProValGlnGlnSerProGlnGlnProAspSerSerSer 20
Db 2614 ACGGCTCCGGGAAAAAGAGAGCGGTAGAGCACTCTCTGTGGAGCCAGACTCTCTCTCG 2673
QY 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyThrGly 40
Db 2674 GGAACCGGAAAGGCGGCGCAGCGCTGCAAGAAAAAGATTGAATTTTGGTCAGACTGGA 2733
QY 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
Db 2734 GACGAGACTCAGTACCTGACCCCGCAGCTCTCGGACAGCCACCGAGCCCTCTCGT 2793
QY 61 ValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGly 80
Db 2794 CTGGAACTAATACGATGGCTACAGGCAAGTGGCGCCCAATGGCAGACAAATAACGAGGC 2853
QY 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
Db 2854 GCCAGCGAGTGGGTAAATCTCGGAATTTGGCAATTCGATTCACATGATGATGGCGAC 2913
QY 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyr 120
Db 2914 AGAGTCATCACCCAGCAGCCGAACTGGGCGCTCCGACCTACCAACAACCCACTCTAC 2973
QY 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer 140
Db 2974 AAACAAATTTCCAGCCAAATCA---GGAGCCTCGAACGACAAATCACTATTGGGTACAGC 3030
QY 141 ThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
Db 3031 ACCCTTGGGGGTATTTTGACTTCAACAGATTCCACTGCCACTTTTACCAGCTGACTGG 3090
QY 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
Db 3091 CAAGACTCATCAACAACAACTGGGGATTCCGACCCCAAGAGACTCAACTTCAAGCTCTTT 3150
QY 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeu 200
Db 3151 AACATTCAAGTCAAAAGAGGTCAACGAGTACGCGTACGAGACGATTTGCCAATACCTT 3210
QY 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
Db 3211 ACCAGCAGGTTCAGTGTCTTACTGACTCGGAGTACAGCTCCCGTACGCTCTCGGCTCG 3270
QY 221 AlaHisGlnGlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGly 240
Db 3271 GCGCATCAAGGATGCTCCCGCGTTCCGAGCAGACGCTCTTCATGTGTGCACAGTATGGA 3330
QY 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysValLeuGlu 260
Db 3331 TACCTCACCCCTGAACAACAGGGAGTCAGGAGTACGAGCGCTCTCTATTCTTACCTCGGAG 3390
QY 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
Db 3391 TACTTTCCTCTCAGATGCTGGTACCGGAACAACTTTTACCTTACGCTTACACTTTTGGAG 3450
```



```
QY 81 AlaAspGlyValGlyAsnAlaSerGlyAenTropHisCysAspSerThrTrpLeuGlyAasp 100
Db 652 GCCGAGGAGTGGGAATTCCTCGGGAATTCGGATTCACATGGATGGCGAC 711
QY 101 ArgValIleThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyr 120
Db 712 AGAGTTATCACCAACAGCAAGAACCTGGGCGCTCCCACTCAATAATCACCTCTAC 771
QY 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer 140
Db 772 AAGCAAAATCTCCAGCGAAATCG---GGAGCCACCAACGACCACTACTCTCGGTACACG 828
QY 141 ThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
Db 829 ACCCTCTGGGGTATTTTACATTTTAAACAGATTCACCTGTCACTTCTCACCAGTCACTGG 888
QY 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnAsnPhelysLeuPhe 180
Db 889 CAGCGACTCATCAACAACCTGGGGATTTAGACCCAGAACTCAATTTCAAGCTCTTC 948
QY 181 AsnIleGlnValLysGluValThrAsnAspGlyValThrThrIleAlaAsnLeu 200
Db 949 AACATCCAACTCAAGGAGTCAACGAGATGATGGAAACCAACCATCGCCCAATACCTT 1008
QY 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
Db 1009 ACCAGCACGGTGCAGGTCTTCACAGACTCTGAGTACCAGCTGCCCTACGTCTCGGTTGG 1068
QY 221 AlaHisGlnGlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGly 240
Db 1069 GCTCACAGGGCTGCTTCGGCGCTTCCAGCAGACGCTCTTCATGATTCCTCAGTACGGC 1128
QY 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValArgSerSerPheTyrCysLeuGlu 260
Db 1129 TACTTGACTCTGAACAAATGGAGCCCAAGCGGTAGGACGTTCTCATTTACTGTCTAGAG 1188
QY 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
Db 1189 TATTTTCCCTCTCAGATGCTGAGGACGGGAACAACCTTCACCTTCAGCTACACTTTTGAA 1248
QY 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
Db 1249 GAGTGCCTTTCCACAGCAGCTACGCGCACAGCAGATCTGGATCGGCTGATGAATCCT 1308
QY 301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
Db 1309 CTCATTGACGAGTACCTGATTACCTGAGCAAACTCAGGGTACAGGTGGAACAACGCGAG 1368
QY 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
Db 1369 CAATCGAGACTGCAGTTCCAGCAAGCTGGGCGCTAGCTCCATGGCTCAGCGAGGCCAAAAC 1428
QY 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
Db 1429 TGGCTACCGGGACCCAGCTACCGACAGCAGCGAATGCTTAAGACGGCTAATGACAAACAAC 1488
QY 361 AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 380
Db 1489 AACAGTGAATTTCTCTGGACTGCGAGCCACCAATATTACTGAATGGAAAGAAATCTCTCG 1548
QY 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet 400
Db 1549 GTCAATCCCGGGCCCCCAATGGCCAGTCAACAGGACGATGAGGAAAAGTATTTCGCCCATG 1608
QY 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
Db 1609 CACGGAATCTCATCTTTGGAAAACAAGGCACAGGAACTACCAATGTGGACATTTGAATCA 1668
QY 421 ValMetIleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
Db 1669 GTGCTTATTACAGCAAGAAAGAAATACAGAACCTAATCTCTGTGGCTACAGAACATAC 1728
QY 441 GlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHis 460
```

```
Db 1729 GGACAGGTGGCCCAACCATCAGAGTCAGAACACACACAGCTTCCTATGGAAGTGGAC 1788
QY 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro 480
Db 1789 AGCCAGGGAATCTTTACCTGGAATGTTGGCAGACCGCGATGTCTATCTTTCAAGGTCCC 1848
QY 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
Db 1849 ATTTGGGCCAANAATCTCTCACACGACGGACACTTTTCATCTCTTCGCTCATGGGAGGC 1908
QY 501 PheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsn 520
Db 1909 TTTGGACTGAAAACACCTCTCCCTCCAGATCCTGATCAAAAACACACCTGTGCCAGCGAAT 1968
QY 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
Db 1969 CCCCGACCACTTTTCACTCTCTGGAAGTTTCCTTCGTTTCATTACCCAGTATTCACCCGGA 2028
QY 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPro 560
Db 2029 CAGTTCAGCTGGGAATAGAGTGGAGCTGCAGAAAGAAACAGCAACGCTGGAACCCA 2088
QY 561 GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAsnValAspPheThrValAspAsn 580
Db 2089 GAAATTCAGTACACCTCCCACTACACAAGTCGGTGAATGTGGAGTTTACCGTGGACGCA 2148
QY 581 AsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db 2149 AACGGTGTATTATTCTGAACCCCGCCCTATTGGCACTCGTTACCTTACCCGGAACCTTG 2205
```

Search completed: February 14, 2004, 22:40:21

Job time : 6271.04 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 14, 2004, 15:18:00 ; Search time 410.871 Seconds
(without alignments)
3935.452 Million cell updates/sec

Title: US-09-807-802A-15

Perfect score: 3251

Sequence: 1 TAPGKRVPESQBPQSSSS.....NNGLYTEPRPGTRYLRLPL 599

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 252756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh
-MODEL=frame+ p2n.model
-Q=/cgn2_1/USPTO spo01/US09807802/runat_11022004_175606_15919/app_query.fasta_1.2389
-DB=N Geneseq 19Jun03 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -HEAFSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09807802 @CGN_1_1_806 @runat_11022004_175606_15919 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|---------------------|
| 1 | 3251 | 100.0 | 1800 | 21 | Adeno-associated v |
| 2 | 3251 | 100.0 | 2211 | 21 | Adeno-associated v |
| 3 | 3251 | 100.0 | 4718 | 21 | Adeno-associated v |
| 4 | 3229 | 99.3 | 4683 | 22 | AAV6 DNA sequence. |
| 5 | 2906 | 89.4 | 1605 | 21 | Adeno-associated v |
| 6 | 2832 | 87.1 | 4722 | 22 | AAV3B DNA sequence |
| 7 | 2801 | 86.2 | 8178 | 21 | Chimeric adeno-ss |
| 8 | 2764.5 | 85.0 | 4072 | 22 | Adenovirus helper |
| 9 | 2764.5 | 85.0 | 4679 | 22 | Adeno-associated v |
| 10 | 2764.5 | 85.0 | 4679 | 22 | AAV2 DNA sequence. |
| 11 | 2764.5 | 85.0 | 4679 | 24 | Adeno-associated v |
| 12 | 2764.5 | 85.0 | 4679 | 25 | ABV76133 |
| 13 | 2764.5 | 85.0 | 7557 | 22 | Adeno-associated v |
| 14 | 2764.5 | 85.0 | 8698 | 22 | AAH26326 |
| 15 | 2722.5 | 83.7 | 4680 | 17 | AAH26325 |
| 16 | 2722.5 | 83.7 | 4680 | 25 | AAH14497 |
| 17 | 2706.5 | 83.3 | 4675 | 22 | Adeno-associated v |
| 18 | 2706.5 | 83.3 | 4675 | 22 | Nucleotide sequenc |
| 19 | 2706.5 | 83.3 | 4675 | 24 | Human adeno-associ |
| 20 | 2706.5 | 83.3 | 4675 | 24 | Human adeno-associ |
| 21 | 2706.5 | 83.3 | 4675 | 24 | ABSG9880 |
| 22 | 2398.5 | 73.8 | 8151 | 21 | Adeno-associated v |
| 23 | 1830.5 | 56.3 | 1800 | 19 | AAV4 VP2 coat prot |
| 24 | 1830.5 | 56.3 | 1800 | 25 | Adeno-associated v |
| 25 | 1830.5 | 56.3 | 2208 | 19 | AAV4 VP1 capsid pr |
| 26 | 1830.5 | 56.3 | 2208 | 25 | Adeno-associated v |
| 27 | 1830.5 | 56.3 | 4767 | 19 | AAV4 genome. Aden |
| 28 | 1830.5 | 56.3 | 4767 | 25 | Adeno-associated v |
| 29 | 1820.5 | 56.0 | 7215 | 21 | Hybrid adeno-associ |
| 30 | 1719 | 52.9 | 2487 | 18 | Duck parvovirus ca |
| 31 | 1701 | 52.3 | 4570 | 16 | Barbary duck parvo |
| 32 | 1700.5 | 52.3 | 2264 | 21 | Adeno associated v |
| 33 | 1700.5 | 52.3 | 2264 | 21 | Adeno associated v |
| 34 | 1700.5 | 52.3 | 2307 | 21 | Adeno associated v |
| 35 | 1700.5 | 52.3 | 4652 | 21 | Adeno associated v |
| 36 | 1700.5 | 52.3 | 4652 | 24 | Adeno-associated v |
| 37 | 1690.5 | 52.0 | 1617 | 19 | AAV4 VP3 coat prot |
| 38 | 1690.5 | 52.0 | 1617 | 25 | Adeno-associated v |
| 39 | 772 | 23.7 | 2271 | 21 | Adeno-associated v |
| 40 | 497.5 | 15.3 | 2343 | 20 | Erythrovirus V9 DN |
| 41 | 497.5 | 15.3 | 4677 | 16 | Human parvovirus g |
| 42 | 497.5 | 15.3 | 5028 | 20 | Genomic DNA sequen |
| 43 | 493 | 15.2 | 2380 | 25 | Human parvovirus B |
| 44 | 493 | 15.2 | 2380 | 25 | Human parvovirus B |
| 45 | 493 | 15.2 | 4678 | 25 | Human parvovirus B |

ALIGNMENTS

RESULT 1
AAD00778
ID AAD00778 standard; DNA; 1800 BP.
XX
AC AAD00778;
XX
DT 08-SEP-2000 (first entry)
XX
DE Adeno-associated virus serotype 1 capsid protein VP2 DNA.
XX
KW Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
KW cap protein; recombinant viral vector; gene delivery; gene therapy;
KW vaccine; transgene; VP2; BS.
XX
OS Adeno associated virus serotype 1.
XX
FH Key Location/Qualifiers

Db 1381 GCTATGGGAGCATTACCTGGCATGGTGGCAAGATAGAGACGTGTACTGCGAGGTCCC 1440
 Qy 481 ileTpaAlaLyLeileProHieTherAepGlyHisPheHisProSerProLeuMetGlyGly 500
 Db 1441 ATTGGGCCAAATTCCTCACACAGATGACACTTTTACCCGCTCTCTTATGGCGGC 1500
 Qy 501 PheGlyLeuLyAsnProProGlnIleLeuIleLeuIleLeuIleLeuIleLeuIleLeuIle 520
 Db 1501 TTTGGACTCAGAAGCCGCTCTCTCAGATCCCTCATCAAAACACACGCTGTTCCTCGAAT 1560
 Qy 521 ProProAlaGluPheSerAlaThrLySLePheAlaSerPheIleThrGlnTySerThrGly 540
 Db 1561 CTCGCGCGAGTTTTCAGTACAAAGTTTGTCTTATTCATCCCACTACTCCACAGA 1620
 Qy 541 GlnValSerValGluIleGluTrpGluLeuGlnLySLeuGlnLySLeuGlnLySLeuGlnPro 560
 Db 1621 CAAAGTGAGTGTGAAATTAATGGAGCTGCAGAAAGAAACAGCAAGCGCTGGAATCCC 1680
 Qy 561 GluValGlnTyThrSerAsnTyAlaLySLeSerAlaAsnValAspPheThrValAspAsn 580
 Db 1681 GAAGTGCAGTACACATCCCAATATGCAAAATCTGCCAACGTTGATTTTACTGTGACAAC 1740
 Qy 581 AnGlyLeuTyThrGluProArgProIleGlyThrArgTyLeuThrArgProLeu 599
 Db 1741 AATGGACTTTTACTGAGCTCGCCCATTTGSCACCCGTTACTTACCGTCCCTG 1797

RESULT 2

AAD00777
 ID AAD00777 standard; DNA; 2211 BP.

XX AAD00777;

DT 08-SEP-2000 (first entry)

DE Adeno-associated virus serotype 1 capsid protein VP1 DNA.

KW Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
 KW cap protein; recombinant viral vector; gene delivery; gene therapy;
 KW vaccine; transgene; VP1; ss.

OS Adeno associated virus serotype 1.

FH Key Location/Qualifiers
 FT CDS 1..2211

FT /*tag= a
 FT /product= "VP1 protein"

PN WO200028061-A2.

XX 18-MAY-2000.

XX 02-NOV-1999; 99WO-US25694.

XX 05-NOV-1998; 98US-0107114.

XX (UYPE-) UNIV PENNSYLVANIA.

XX Wilson JM, Xiao W;

XX WPI; 2000-376571/32.

XX P-PSDB; AAY71167.

XX Novel adeno-associated virus serotype 1 polynucleotide useful for
 PT preparation of medicament for delivery of a transgene to a host -

XX Claim 10; Page 83-87; 108pp; English.

XX The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA
 CC which is characterised by two inverted terminal repeats (ITR) and open
 CC reading frames for rep and capsid (cap) proteins. The rep reading frame
 CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
 CC reading frame encodes three structural proteins, VP1, VP2 and VP3.
 CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap

CC coding regions, are useful in production of recombinant viral vectors
 CC for gene delivery. These vectors can be used as gene therapy
 CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
 CC not induce the formation of neutralising antibodies specific to any
 CC serotype of AAV hence is useful for transforming host cells, and in the
 CC preparation of a medicament for the delivery of transgene to a host.
 CC The present sequence is an AAV-1 DNA encoding a cap protein VP1 which is
 CC useful in the production of recombinant viral vector for gene delivery.

SQ Sequence 2211 BP; 577 A; 652 C; 538 G; 444 T; 0 other;

Alignment Scores:
 Pred. No.: 1,44e-259 Length: 2211
 Score: 3251.00 Matches: 599
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 Dbs: 21 Gaps: 0

US-09-807-802A-15 (1-599) x AAD00777 (1-2211)

Qy 1 ThrAlaProGlyLySLeysArgProValGluGlnSerProGlnGluProAspSerSerSer 20
 Db 412 ACGGCTCTCTGGAAGAAAGCTCCGCTAGAGCAGTCGCCACAAAGAGCCAGACTCTCTCTCG 471
 Qy 21 GlyIleGlyLySThrGlyGlnGlnProAlaLySLeysArgLeuAsnPheGlyGlnThrGly 40
 Db 472 GGCATCGGCAAGACAGCCAGCCGCTCTGAGAGAACTGCTGAGAAAGACTCAATTTTGGTCAGACTGGC 531
 Qy 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
 Db 532 GACTCAGAGTCAGTCCCGCATCCAAACCTCTCGAGAACTCTCCAGCAACCCCGCTGCT 591
 Qy 61 ValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGly 80
 Db 592 GTGGGACCTACTACAAATGGCTTCAGGCGGTGGCGCACCACCAATGGCAGACAATAACGAAGGC 651
 Qy 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTirpleuGlyAsp 100
 Db 652 GCCAGCGAGTGGGTAAATGCCCTCAGGAAATTTGGGATTCGATTCACATGCTGGGCGCAC 711
 Qy 101 ArgValIleThrThrSerThrArgThrTirAlaLeuProThrTyAsnAsnHisLeuTy 120
 Db 712 AGAGTCATCACACAGCAGCCCGCCTCTGGGCTTGGCCACTACATATACCACTCTTAC 771
 Qy 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyPheGlyTySer 140
 Db 772 AAGCAATCTCCAGTGTCTTCAACGGGGCCAGCAACGACCAACCACTACTTCGGGTACAGC 831
 Qy 141 ThrProTrpGlyTyPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
 Db 832 ACCCCCTGGGGGTATTTTGATTTCAACAGATTCACACTGCCACTTTTCCACCGTACTGG 891
 Qy 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLySArgLeuAsnPheLySLeuPhe 180
 Db 892 CAGCGACTCATCAACACACATTTGGGGATTCGGGCCCAAGAGACTCAACTTCAAACTCTTC 951
 Qy 181 AsnIleGlnValLySLeuValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeu 200
 Db 952 AACATCCAAAGTCAAGGAGGTCAACGCAATGATGGCGTCACACCAACCATCGCTAATAACCTT 1011
 Qy 201 ThrSerThrValGlnValPheSerAspSerGluTyGlnLeuProTyTyValLeuGlySer 220
 Db 1012 ACCAGCAGCGTTCAGGTCTTCTCGGACTCGGAGTACCACTTCCGTCGTCTCTCGGCTCT 1071
 Qy 221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyGly 240
 Db 1072 GCGCACCAAGGCTGCT 1131
 Qy 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyCysLeuGlu 260
 Db 1132 TACCTGACGCTCAACAATGGCAGCAAGCGCTGGGACGTTTCATCTCTTTTACTGCTCGGAA 1191

QY 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
DB |||||||
DB 1192 TATTTCCTTCCTCAGATGCTGAGACGGGCAACACTTTACCTTCAGCTACACCTTTGAG 1251
QY 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
DB |||||||
DB 1252 GAAGTGCTTTCCACAGCAGCTACGGCGCACAGCAGCGCTGGACCGCTGATGAATCT 1311
QY 301 LeuLeuAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
DB |||||||
DB 1312 CTCATCGACCAATACCTGATTTACCTGAACAGAACTCAAAATCAGTCGGAAAGTCCCAA 1371
QY 321 AsnLeuAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
DB |||||||
DB 1372 AACAGAGCTTGCTGTTTACCGTGGGTCTCCAGCTGGGATGTCTGTTCAGGCCCAAAAC 1431
QY 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
DB |||||||
DB 1432 TGGCTACCTGGACCCCTGTTATCGGCAGCAGCGCGTTCATAAACAACAAACAGACCAAC 1491
QY 361 AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 380
DB |||||||
DB 1492 AACAGCAATTTTACCTGGAGCTGGTCTCAAAATATAAATCACTCAATGGCGTGAATCCATC 1551
QY 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPheProMet 400
DB |||||||
DB 1552 ATCAACCTGGCACTGCTATGGCTCACAAACAGACGACGAGCAAGTCTCTTCCCATG 1611
QY 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
DB |||||||
DB 1612 AGCGGTGTCATGATTTTGGAAAGAGAGCGCGGAGCTTCAACACACTGCATTTGGACAAT 1671
QY 421 ValMetIleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
DB |||||||
DB 1672 GTCATGATTACACAGCAAGAGGAAATTAAGCCACTAACCCCTGTGGCCACCGGAAAGATT 1731
QY 441 GlyThrValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHis 460
DB |||||||
DB 1732 GGGACGCTGGCATCTCAATTTCCAGAGCAGCAGCAGACCCCTGCGACCGGAGATGTCAT 1791
QY 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro 480
DB |||||||
DB 1792 GCTATGGAGCATTTACCTGGCATGGTGTGGCAAGATAGACGCTGTACCTGCAGGGTCCC 1851
QY 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
DB |||||||
DB 1852 ATTTGGGCCAAAATTCCTCACAGATGGACATTTTACCCCGTCTCTCTTATGGCGGC 1911
QY 501 PheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsn 520
DB |||||||
DB 1912 TTTGGACTCAAGAACCCGCTCTCAGATCTCATCAAAACACGCTGTCTCTGCGAAT 1971
QY 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
DB |||||||
DB 1972 CCTCCGCGGAGTTCCTCAGCTACAAAGTTTGTCTTCATTCATCCCAATACTCCACAGGA 2031
QY 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPro 560
DB |||||||
DB 2032 CAAGTGAGTGTGGAAATTAATGGAGCTCGAGAAAGAAACAGCAAGCGCTGGATCCC 2091
QY 561 GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsn 580
DB |||||||
DB 2092 GAAGTGAGTACACATCCAAATTAATGCAAAATCTGCCAACCTGATTTTACTGTGGCAAC 2151
QY 581 AsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
DB |||||||
DB 2152 AATGGACTTTTACTAGCTCGCCCAATGGCACCCGTTACTTACCCGTCCTCCCTG 2208

RESULT 3
AAD00772
ID AAD00772 standard; DNA; 4718 BP.
XX
AC AAD00772;

XX 08-SEP-2000 (first entry)
DT
XX Adeno-associated virus serotype 1 DNA.
DE
XX Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
KW cap protein; recombinant viral vector; gene delivery; gene therapy;
KW vaccine; transgene; ss.
XX
OS Adeno associated virus serotype 1.
XX
FH Key Location/Qualifiers
FT repeat_unit 1..143
FT /tag= a
FT /label= 5' ITR
FT /note= "Inverted terminal repeat which is capable
FT of forming T-shaped hairpin structure"
FT 89..110
FT /tag= b
FT /bound_moiety= "Rep protein"
FT 124..125
FT /tag= c
FT /note= "Terminal residue site (TRS)"
FT 219..226
FT /tag= d
FT /bound_moiety= "USF"
FT /note= "E box"
FT 236..239
FT /tag= e
FT /label= P5_promoter
FT 237..245
FT /tag= f
FT /bound_moiety= "YV1 factor"
FT 270..275
FT /tag= g
FT /label= P5_TATA-Box
FT 299..306
FT /tag= h
FT /note= "YV1/p5 RNA"
FT 335..2206
FT /tag= i
FT /product= "Rep 78"
FT /function= "regulates replication and integration
FT of AAV DNA into host cell's chromosome"
FT 335..2272
FT /tag= j
FT /product= "Rep 68"
FT /function= "regulates replication and integration
FT of AAV DNA into host cell's chromosome"
FT /note= "The coding region is interrupted by intron"
FT 1924..2220
FT /tag= k
FT /note= "This region interrupts the coding sequence
FT of Rep 68 and Rep 40"
FT 857..862
FT /tag= l
FT /label= P19_TATA_Box
FT 882..883
FT /tag= m
FT /note= "P19 RNA"
FT 1007..2206
FT /tag= n
FT /product= "Rep 52"
FT /function= "regulates replication and integration
FT of AAV DNA into host cell's chromosome"
FT 1007..2272
FT /tag= o
FT /product= "Rep 40"
FT /function= "regulates replication and integration
FT of AAV DNA into host cell's chromosome"
FT /note= "The coding region is interrupted by intron"
FT 1836..1841
FT /tag= p
FT

```

FT misc_feature /label= P40_TATA-BOX
FT 1875..1876
FT /*tag= q
FT /note= "P40 RNA"
FT CDS
FT 2223..4433
FT /*tag= r
FT /product= "vp1 protein"
FT /note= "Capsid protein"
FT CDS
FT 2634..4433
FT /*tag= s
FT /product= "vp2 protein"
FT /note= "Capsid protein"
FT partial
FT 2829..4433
FT /*tag= t
FT /product= "vp3 protein"
FT /note= "Capsid protein"
FT CDS
FT 4447..4452
FT /*tag= u
FT polyA_signal 4576..4718
FT repeat_unit /*tag= v
FT /label= 3' ITR
FT /note= "Inverted terminal repeat which is capable
FT of forming T-shaped hairpin structure"
XX
XX WO200028061-A2.
XX
XX 18-MAY-2000.
XX
XX 02-NOV-1999; 99WO-US25694.
XX
XX 05-NOV-1998; 98US-0107114.
XX
XX (UYPE-) UNIV PENNSYLVANIA.
XX
XX Wilson JM, Xiao W;
XX
XX WPI; 2000-376571/32.
XX P-PSDB; AAY71161, AAY71164, AAY71165, AAY71166, AAY71167, AAY71168,
XX AAY71169.
XX
XX Novel adeno-associated virus serotype 1 polynucleotide useful for
XX preparation of medicament for delivery of a transgene to a host
XX
XX Claim 1; Fig 1; 108pp; English.
XX
XX The present sequence is an adeno-associated virus serotype 1 (AAV-1)
XX DNA characterised by two inverted terminal repeats (ITR) and open
XX reading frames for rep and capsid (cap) proteins. The rep reading frame
XX encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
XX reading frame encodes three structural proteins, VP1, VP2 and VP3.
XX The AAV-1 sequence or its fragments particularly ITRs, rep and cap
XX coding regions, are useful in production of recombinant viral vectors
XX for gene delivery. These vectors can be used as gene therapy
XX vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
XX not induce the formation of neutralising antibodies specific to any
XX serotype of AAV hence is useful for transforming host cells, and in the
XX preparation of a medicament for the delivery of transgene to a host.
XX
XX Sequence 4718 BP; 1121 A; 1393 C; 1273 G; 931 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 4.02e-259 Length: 4718
XX Score: 3251.00 Matches: 599
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 21 Gaps: 0
XX
XX US-09-807-802a-15 (1-599) x AAD000772 (1-4718)
XX
XX 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSer 20
XX |||||||||||||||||||||||||||||||||||||||||||||||||||||||

```

```

Db 2634 ACGGCTCCTGGAAAGAAACGTCGGTAGACGATCGCCACACAGACCGACCTCTCTCTCG 2693
Qy 21 GlyIleGlyLysThrGlyGlnProAlaLysLysArgLeuAenPheGlyGlnThrGly 40
Db 2694 GGATCGGCAAGACAGCGCCAGCAGCCCGCTAAAAAGAGACTCAATTTTGGTCAGACTGGC 2753
Qy 41 AspSerGluSerValProAspProGlnProLeuGlyGluProAlaThrProAlaAla 60
Db 2754 GACTCAGAGTCAGTCCCGGATCCCAACCTCTCCGAGAACCTCCAGCAACCCCCGCTGCT 2813
Qy 61 ValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAenAsnGluGly 80
Db 2814 GTGGACCTACTACATGGCTTCAGCGGTCGCGCAGCAACCAATGGCAGACAATAACGAAGGC 2873
Qy 81 AlaAspGlyValGlyAenAlaSerGlyAenTrpHisCysAspSerThrTrpLeuGlyAsp 100
Db 2874 GCCGACGAGTGGGTAATGCTTCAGGAAATGGCAATTCGCAATCCACATGGCTGGCGAC 2933
Qy 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTyrAenAsnHisLeuTyr 120
Db 2934 AGAGTCATCACCCACGACCCCGCACCTGGGCTTCGCCACCTACCAATTAACACCTCTAC 2993
Qy 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAenAspAenHisTyrPheGlyTyrSer 140
Db 2994 AAGCAATCTCCAGTGTCTCAACGGGGCCAGCAACGACCACTACTTCCGCTACAGC 3053
Qy 141 ThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
Db 3054 ACCCCCTGGGGTATTTGATTTCAACAGATTCACACTGCCACTTTTCAACACGACTGG 3113
Qy 161 GlnArgLeuIleAenAenAenTrpGlyPheArgProLysArgLeuAenPheLysLeuPhe 180
Db 3114 CAGGACTCATCAACAACAATTTGGGGATTCGGGCCCAAGAGACTCAACTTCAAACTCTTC 3173
Qy 181 AenIleGlnValLysGluValThrAenAspGlyValThrThrIleAlaAenAsnLeu 200
Db 3174 AACATCCAGTCAAGGAGGTCAACGCGATGATGGCGTCACACCATCGCTAATAACCTT 3233
Qy 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
Db 3234 ACCGACGCGTTCAGTCTCTCGACTCGGAGTACCAAGCTTCGCTACGCTCTCTCGGCTCT 3293
Qy 221 AlaHisGlnGlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGly 240
Db 3294 GCGCACCAAGGCTGCTCCCTCCGTTCCGCGGAGCGTTCATGATTCGCAATACGCGC 3353
Qy 241 TyrLeuThrLeuAenAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
Db 3354 TACCTGACGCTCAACAATGGCAGCCAGCCGCTGGGACGTTTCATCTTTACTCCCTGGA 3413
Qy 261 TyrPheProSerGlnMetLeuArgThrGlyAenAenPheThrPheSerTyrThrPheGlu 280
Db 3414 TATTTCCCTTCTCAGATGCTGAGAACCGGCAACAACCTTTACCTTCAGCTACACCTTGA 3473
Qy 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAenArgLeuMetAenPro 300
Db 3474 GAAGTGCCTTTCCACAGCAGCTACCGCAGCAGCCAGCAGCTGGACCGGCTGATGATCTCT 3533
Qy 301 LeuIleAspGlnTyrLeuTyrTyrLeuAenArgThrGlnAenGlnSerGlySerAlaGln 320
Db 3534 CTCATCGACCAATACCTGTATTACCTGAACAGACTCAAAATCAGTCGGAAGTGCCTCAA 3593
Qy 321 AenLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAen 340
Db 3594 AACAGGACTTGTGCTTTAGCGTGGGTCTCCAGCTGGCATGTCTGTTCAGCCCAAAAC 3653
Qy 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAenAen 360
Db 3654 TGGCTACCTGGACCTGTTATCGGCGAGCGCGCTTTCTAAACAAAACAGACACCAAC 3713
Qy 361 AenSerAenPheThrTrpThrGlyValaSerLysTyrAenLeuAenGlyArgGluSerIle 380
Db 3714 AACACCAATTTACCTGGACTGGTGTCTCAAAATATACCTCAATGGGCGTGAATCCATC 3773

```



```
QY 381 lIeAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet 400
Db 3774 ATCAACCTTGGGCACTGCTATGGCTTACACAAAGACGACGAAGACAAGTTCTTTCCCATG 3833
QY 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
Db 3834 AGCGGTGTCATGATTTTGGAAAAGAGAGCGCGGAGCTTCAACACATGCTATGGACAAT 3893
QY 421 ValMetIleThrAspGluGluLysAlaThrAsnProValAlaThrGluArgPhe 440
Db 3894 GTCATGATTACAGCAGAGGAAATTAAGCCACTTAACCTTGTGGCCACCGAAAGATTT 3953
QY 441 GlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHis 460
Db 3954 GGGACCGTGGCAGTCATTTCCAGACAGCAGCAGACCCCTGGCCGCGAGATGTGCAT 4013
QY 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTrpLeuGlnGlyPro 480
Db 4014 GCTATGGAGCATTACCTGGCATGGTGTGGCAAGATAGAGAGCTGTATCCTGCGAGGTCCC 4073
QY 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
Db 4074 ATTGGGCGCAAAATTCCTCACACAGATGGACACTTTTCAACCCGCTCTCTTATGGCGGC 4133
QY 501 PheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsn 520
Db 4134 TTGGACTCAGAACCGGCTCTCAGATCCTCATCAAAAACACGCGCTGTCTCTGCGAAT 4193
QY 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTrpSerThrGly 540
Db 4194 CTTCCGGCGAGTTTTCAGCTACAAAGTTTGTCTTCTTATCATCCCAATACTCCACAGGA 4253
QY 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPro 560
Db 4254 CAAGTGAAGTGGAAATTTGAATGGAGCTGCAGAAAGAAACAGCAAGCGCTGGAATCCC 4313
QY 561 GluValGlnTrpThrSerAsnTrpAlaLysSerAlaAsnValAspPheThrValAspAsn 580
Db 4314 GAAGTGCAGTACATCCANTTATGCAAAATCTGCCAAGCTGATTTTACTGTGGACAAC 4373
QY 581 AsnGlyLeuTrpThrGluProArgProIleGlyThrArgTrpLeuThrArgProLeu 599
Db 4374 AATGGACTTTATCTAGCCCTCGCCCATTTGGCACCCTTACCTTACCCGTCGCCCTG 4430

RESULT 4
AAF23749 standard; DNA; 4683 BP.
XX
AC AAF23749;
XX
DT 28-MAR-2001 (first entry)
XX
DE AAV6 DNA sequence.
XX
KW AAV6; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
KW atherosclerosis; sickle cell anaemia; thalassemia;
KW blood clotting disorder; diabetes; ss.
XX
OS Adeno associated virus.
XX
PN US6156303-A.
XX
PD 05-DEC-2000.
XX
PF 11-JUN-1997; 97US-0873168.
XX
PR 11-JUN-1997; 97US-0873168.
XX
PA (UNIW ) UNIV WASHINGTON.
XX
PI Russell DW, Rutledge EA;
XX
```

DR WPI; 2001-060164/07.

XX Adeno-associated virus serotype 6 and viral vector derived from it for
PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
PT syndrome, sickle cell anemia, thalassemia and diabetes
XX
PS Claim 1; Fig 1; 50pp; English.
XX
CC The present invention relates to adeno-associated virus serotypes. The
CC present sequence is the DNA sequence of one such serotype (AAV6). AAV6
CC can be used to construct AAV viral vectors for use in gene therapy for a
CC range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
CC sickle cell anaemia, thalassemia, blood clotting disorders and diabetes.
CC The AAV viral vectors have increased transduction efficiency of a
CC particular host cell as the AAV virion containing the AAV vector genome
CC can be modified to express a capsid protein of an AAV serotype that
CC transduces the selected host cell.

XX
SQ Sequence 4683 BP; 1114 A; 1363 C; 1277 G; 929 T; 0 other;

Alignment Scores:
Pred. No.: 2,65e-257 Length: 4683
Score: 3229.00 Matches: 594
Percent Similarity: 99.67% Conservations: 3
Best Local Similarity: 99.17% Mismatches: 2
Query Match: 99.32% Indels: 0
DB: 22 Gaps: 0

US-09-807-802A-15 (1-599) x AAF23749 (1-4683)

QY 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSer 20
Db 2619 ACGGCTCTCGAAAGAAACGTCGCGTAGAGCAGTCGCCACAGAGCCAGACTCTCTCTCG 2678
QY 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLysAsnPheGlyGlnThrGly 40
Db 2679 GGCATTTGGCAGACAGCGCCAGCAGCCCGCTTAAAGAGACTCAATTTTGGTCAGACTGGC 2738
QY 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
Db 2739 GACTCAGAGTCAGTCCCCGACCCACCACTCTCGGAGAACCTCCAGCAACCCCGCTGT 2798
QY 61 ValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGlyGly 80
Db 2799 GTGGGACCTTACTACATGCGCTTCAGCGCGTGGCGCACCACCAATGGCAGACAATACGAGGC 2858
QY 81 AlaAspGlyValGlyAsnAlaSerGlyValAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
Db 2859 GCCGACGGAGTGGGTAAATGCTTCAGGAATTTGGCATTTGCCATTCACATGGCTGGCGAC 2918
QY 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTrpAsnAsnHisLeuTrp 120
Db 2919 AGAGTCATCCACCACCCAGCAGCCGAACTGGGCTTTGCCACCTATAACAACACCTCTAC 2978
QY 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTrpPheGlyTrpSer 140
Db 2979 AAGCAAAATCTCCAGTGTCTTCAACGGGGCGCCAGCAACACCACTACTTCCGGCTACAGC 3038
QY 141 ThrProTrpGlyTrpPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
Db 3039 ACCCTCTGGGGTATTTTGGATTTCAACAGATTTCCACTGCCATTTCTCCACCGTACTGG 3098
QY 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
Db 3099 CAGCGACTCATCAACAACAAATTTGGGATTTCCGCGCCCAAGAGACTCAACTTCAAGCTCTTC 3158
QY 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeu 200
Db 3159 AACATCCAAAGTCAAGAGAGGTCCAGCAAGATGATGGCGTCCAGCACCATCTCCATTAACCTT 3218
QY 201 ThrSerThrValGlnValPheSerAspSerGluTrpGlnLeuProTrpValLeuGlySer 220
Db 3219 ACCAGCACGGTTCNAGTCTTCTCGGACTCGGAGTACCAGTTCGCCGTACCTCTCGGCTCT 3278

```
Qy 221 AlaHisGlnGlyCysLeuProPheProPheProAlaHisValPheMetIleProGlnTyrGly 240
Db 3279 GCGACACGAGGCTGCTCCCTCCGTTCCGCGGAGCGTGTTCATGATTCGCGACGTGCGC 3338
Qy 241 TyrLeuThrLeuAenAenGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
Db 3339 TACTTACGCTCAACATGGCAGCAGCAGCAGTGGCAGTCTCTCTTTACTGCTCGAA 3398
Qy 261 TyrPheProSerGlnMetLeuArgThrGlyAenAenPheThrPheSerTyrThrPheGlu 280
Db 3399 TATTTCATCGCAGATGCTGAGAACGGGCAATAACTTTTACCTTCAGCTACACCTTCGAG 3458
Qy 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAenAenAenAenAenAen 300
Db 3459 GACGTGCTTTTCCACAGCAGCTAGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3518
Qy 301 LeuIleAepGlnTyrLeuTyrTyrLeuAenAenArgThrGlnAenGlnSerGlySerAlaGln 320
Db 3519 CTATCGACCACTGATCTGATTAACCTGACAGAACTCAGATCAGTCCGGAGTGGCCNA 3578
Qy 321 AenLysAepLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAen 340
Db 3579 AACAAAGGACTTGTCTTTAGCGGGGTCTCCAGCTGGCATGTCTGTTCAGCCCAAAAC 3638
Qy 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAenAen 360
Db 3639 TGGCTACCTGGACCTGTATCCGCGCAGCAGCGCGTTCATAAACAACAAACAGACAAAC 3698
Qy 361 AenSerAenPheThrTyrThrGlyValSerLysTyrAenLeuAenGlyArgGluSerIle 380
Db 3699 AACAGCAACTTACTTGGACTGGTGTCTCAAAATATAACCTTAAATGGCGGTGATCTATA 3758
Qy 381 IleAenProGlyThrAlaMetAlaSerHisLysAepAepGluAepLysPhePheProMet 400
Db 3759 ATCAACCTGGCACTGTATGGCTTATGGCTTCAAAATATAACCTTAAATGGCGGTGATCTATA 3818
Qy 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAenThrAlaLeuAenAen 420
Db 3819 AGCGTGTCATGATTTTGGAAAGAGAGAGCGCGGAGCTTCAACACCTGCATTTGGACAT 3878
Qy 421 ValMetIleThrAspGluGluLeuLysAlaThrAenProValAlaThrGluArgPhe 440
Db 3879 GTCATGATCACAGCAGAGAGAAATCAAGCCACTAACCCCGTGGCCACCGAAGATTT 3938
Qy 441 GlyThrValAlaValAenPheClnSerSerSerThrAepProAlaThrGlyAspValHis 460
Db 3939 GGGACTGTGGCAGTCAATCTCCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3998
Qy 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAepArgAepValTyrLeuGlnGlyPro 480
Db 3999 GTTATGGAGCTTACCTGGATGTGTGGCAAGACAGACGATATCTCGAGGTCT 4058
Qy 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
Db 4059 ATTTGGGCCAAAATTCCTCACAGGATGGACACTTTCACCCGTCTCTCTCATGGCGGC 4118
Qy 501 PheGlyLeuLysAenProProGlnIleLeuLysAenThrProValProAlaAen 520
Db 4119 TTTGACTTAAAGCAGCCGCTCTCAGATCTCTCATCAAAAACACGCTGTCTTCGCGAAT 4178
Qy 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
Db 4179 CTTCCGCGCAGATTTTGGCTACAAAGTTTGGTTTCTTATTCATCACCAGTATTCACAGGA 4238
Qy 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAenSerLysArgTyrAenPro 560
Db 4239 CAAGTGAGCGTGAGATTGAATGGAGCTGCAGAAAGAAACAGCAACCGTGGGAATCCC 4298
Qy 561 GluValGlnTyrThrSerAenTyrAlaLysSerAlaAenValAepPheThrValAenAen 580
Db 4299 GAAGTGCAGTATACATCTAACTATGCAAAATCTGCCAAGCTTGATTTTCACTGTGGACAAC 4358
```

```
Qy 581 AenGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db 4359 AATGGACTTTATTAAGCTGAGCTCGCCCATTTGGCACCCTTACCTACCCCTCCCTG 4415
RESULT 5
AAD00779
ID AAD00779 standard; DNA; 1605 BP.
XX AAD00779;
AC AAD00779;
XX 08-SEP-2000 (first entry)
XX Adeno-associated virus serotype 1 capsid protein VP3 DNA.
DE Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
XX cap protein; recombinant viral vector; gene delivery; gene therapy;
XX vaccine; transgene; VP3; ss.
XX Adeno associated virus serotype 1.
XX Key Location/Qualifiers
XX CDS 1..1605
XX FT /*tag= a
XX FT /product= "VP3 protein"
XX WC200028061-A2.
XX 18-MAY-2000.
XX 02-NOV-1999; 99WO-US25694.
XX 05-NOV-1998; 98US-0107114.
XX (UNP-) UNIV PENNSYLVANIA.
XX Wilson JW, Xiao W;
XX PI
XX WPI; 2000-376571/32.
XX P-PSDB; AAY71169.
XX Novel adeno-associated virus serotype 1 polynucleotide useful for
XX preparation of medicament for delivery of a transgene to a host -
XX Claim 10; Page 96-99; 108pp; English.
XX The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA
XX which is characterised by two inverted terminal repeats (ITR) and open
XX reading frames for rep and capsid (cap) proteins. The rep reading frame
XX encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
XX reading frame encodes three structural proteins, VP1, VP2 and VP3.
XX The AAV-1 sequence or its fragments particularly ITRs, rep and cap
XX coding regions, are useful in production of recombinant viral vectors
XX for gene delivery. These vectors can be used as gene therapy
XX vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
XX not induce the formation of neutralising antibodies specific to any
XX serotype of AAV hence is useful for transforming host cells, and in the
XX preparation of a medicament for the delivery of transgene to a host.
XX The present sequence is an AAV-1 DNA encoding a cap protein VP3 which is
XX useful in the production of recombinant viral vector for gene delivery.
XX Sequence 1605 BP; 431 A; 466 C; 361 G; 347 T; 0 other;
Alignment Scores:
Pred. No.: 3.64e-231 Length: 1605
Score: 2906.00 Matches: 534
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 89.39% Indels: 0
DB: 21 Gaps: 0
US-09-807-802A-15 (1-599) x AAD00779 (1-1605)
Qy 66 MetAlaSerGlyGlyAlaProMetAlaAenAenAenGluGlyAlaAepGlyValGly 85
```

|||||
1 ATGGCTTCAGGCGGTGGCGCCACCAATGGCAGACATATACGAAGGCGCGCAGCTGGGT 60
86 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 105
61 AATGCTTCAGGAAATTTGGCATTTGCCATTCACATGCTGGCGGACAGAGTTCATACCCACC 120
106 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSer 125
121 AGCACCCGCACTGGGCCCTTGGCCACCTACATATACCACTCTACAGCAATCTCCAGT 180
126 AlaSerThrGlyAlaSerAsnAspAsnHisIleThrPheGlyTyrSerThrProTrpGlyTyr 145
181 GCTTCAACGGGCGCCAGCAACGACACCACTACTTCGGCTACAGCACCCCTGGGGGTAT 240
146 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgIleLeu 165
241 TTGATTTTCAACAGATTCCTCACTGCCACTTTTACCACGCTGATGGCAGCGACTCATCAAC 300
166 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysPheAsnIleGlnValLys 185
301 AACAAATGGGATTCGGGCCCAAGAGACTCAACTTCAACTCTTCAACATCCCAAGTCAAG 360
186 GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 205
361 GAGGTCAACGCAATGATGGCGTCACCAACCATCGCTAATAACTTTACCAGCAGCGTTCAA 420
206 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 225
421 GTCTTCGAGCTCGGAGTACCAAGCTTCGGTACGCTCTCGGCTCTCGCACCAGGCGCTGC 480
226 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 245
481 CTCCTTCGTTCCCGCGGAGCGTGTTCATGATTCGCAATACGGCTACCTGACGCTCAAC 540
246 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 265
541 AATGGCAGCAACGCGTGGGAGCTTCATCTTTACTGCTGGAATATTTCCCTTCTCAG 600
266 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 285
601 ATGCTGAGAACGGGCAACAACTTTACTCTTCAGCTACACCTTTGAGGAAGTGGCTTTCCAC 660
286 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 305
661 AGCAGCTACGGGCACAGCCAGAGCTGGACCGGCTGATGATCTCTCATCGCAATATAC 720
306 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 325
721 CTGTATTACCTGAACAGAACTCAAAATCAGTCGGGAAGTGGCCAAACCAAGGACTTGTCTG 780
326 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 345
781 TTTAGCGGTGGGTCTCCAGCTGGCATGTCTGTTTCAGCCCAAAACCTGGCTACTCGTACCC 840
346 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 365
841 TGTATTCGGCAGCAGCGCGTTCCTAATAAACAACAACAGACACCAACAGCAATTTTACC 900
366 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr 385
901 TGGACTGGTGTCTTCAAAATATAACCTCAATGGCGGTGAATCCATCATCAACCTGGCACT 960
386 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 405
961 GCTATGGCTTCACACAAAGACGACGAGCAAGTCTTTCCTCCATGAGCGGTGTCATGAT 1020
406 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 425
1021 TTTGGAAAGAGAGCGCCGAGCTTCAACACTTGCATTCGACATGCTCATGATTACAGAC 1080
426 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 445

1081 GAAGAGGAAATTAAGCCACTAACCCCTGTGTGCCACCGAAAGATTGGGACCGTGGCAGTC 1140
446 AsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 465
1141 AATTTCCAGAGCAGACAGACAGCCCTTGGACCCGAGATGTGCATGTCTATGGGAGCATT 1200
466 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 485
1201 CCTGGCATGGTGTGGCAAGATAGAGACGTGTACCTGAGGGTCCCATTTGGGCCCAAAAT 1260
486 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 505
1261 CCTCACAGATGACACTTTTACCCGCTCTCTCTTATGGCGGCTTTGGACTCAAGAAC 1320
506 ProProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 525
1321 CCGCTCTCTCAGATCTCTCATCAAAAACAGCGCTGTCTCTGGAATCTCTCCGGCGAGTTT 1380
526 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 545
1381 TCAGCTACAAAGTTTGTCTTCAATCATCCCAATACTCCACAGGACCAAGTGTGTGGAA 1440
546 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrThr 565
1441 ATTGAATGGGAGCTGCAGAAAGAAACAGCAGCGCTGGAATCCCGAAGTGCAGTACACA 1500
566 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 585
1501 TCCAAATATGCAAAATCTGCCAAGTTGATTTTACTGTGGACAAACATGGACTTTATACT 1560
586 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
1561 GAGCTCGCCCAATGGCACCCTTACCTTACCCGTCGCCCTG 1602

RESULT 6
AAF23748
ID AAF23748 standard; DNA; 4722 BP.
XX
AC AAF23748;
XX
DT 28-MAR-2001 (first entry)
XX
DE AAV3B DNA sequence.
XX
KW AAV3B; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
KW atherosclerosis; sickle cell anaemia; thalassaemia;
KW blood clotting disorder; diabetes; ss.
XX
OS Adeno associated virus.
XX
PN US6156303-A.
XX
PD 05-DEC-2000.
XX
PF 11-JUN-1997; 97US-0873168.
XX
PR 11-JUN-1997; 97US-0873168.
PA (UNITW) UNIV WASHINGTON.
XX
PI Russell DW, Rutledge EA;
XX
DR WPI; 2001-060164/07.
XX
PT Adeno-associated virus serotype 6 and viral vector derived from it for
PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
PT syndrome, sickle cell anemia, thalassemia and diabetes
XX
PS Example 2; Fig 1; 50pp; English.
XX
CC The present invention relates to adeno-associated virus serotypes. The
CC present sequence is the DNA sequence of one such serotype (AAV3B). AAV3B
CC can be used to construct AAV viral vectors for use in gene therapy for a

XX DE Chimeric adeno-associated virus 2/3 capsid encoding helper plasmid.
 XX KW Adeno-associated virus; AAV2; AAV3; chimeric; gene therapy; plasmid;
 KW recombinant parvoviral vector; cellular tropism; rep gene; cap protein;
 KW capsid; gene delivery; VP1; VP2; VP3; ss.
 XX OS Chimeric - Adeno associated virus serotype 2.
 OS Chimeric - Adeno associated virus serotype 3.
 XX FH Location/Qualifiers
 FT CDS 251..2180
 FT /tag= a
 FT /product= "Rep 68"
 FT /notes= "The specification states that this region
 FT encodes Rep 68 protein of AAV2. However, the stop
 FT position given in the specification for this CDS does
 FT not correspond to stop codon"
 FT 251..2114
 FT /tag= b
 FT /product= "Rep 78"
 FT /notes= "The specification states that this region
 FT encodes Rep 78 protein of AAV2. However, the stop
 FT position given in the specification for this CDS does
 FT not correspond to stop codon"
 FT 923..2114
 FT /tag= c
 FT /product= "Rep 52"
 FT /notes= "The specification states that this region
 FT encodes Rep 52 protein of AAV2. However, the stop
 FT position given in the specification for this CDS does
 FT not correspond to stop codon"
 FT 923..2180
 FT /tag= d
 FT /product= "Rep 40"
 FT /notes= "The specification states that this region
 FT encodes Rep 40 protein of AAV2. However, the stop
 FT position given in the specification for this CDS does
 FT not correspond to stop codon"
 FT 2133..4342
 FT /tag= e
 FT /product= "VP1 cap protein"
 FT /notes= "The specification states that this region
 FT encodes VP1 protein of AAV2. However, the stop
 FT position given in the specification for this CDS does
 FT not correspond to stop codon"
 FT 2544..4342
 FT /tag= f
 FT /product= "VP2 cap protein"
 FT /notes= "The specification states that this region
 FT encodes VP2 protein of AAV2. However, the positions
 FT given in the specification for this CDS do not correspond
 FT to start or stop codons"
 FT 2739..4342
 FT /tag= g
 FT /product= "VP3 cap protein"
 FT /notes= "The specification states that this region
 FT encodes VP3 protein of AAV2 in which loops 2-4 are
 FT replaced with the corresponding region from AAV3.
 FT However, the stop position given in the specification
 FT for this CDS does not correspond to stop codon"
 FT 3184..4092
 FT misc_feature
 FT /tag= h
 FT /note= "AAV3 sequence"
 XX PN W0200028004-A1.
 XX PD 18-MAY-2000.
 XX PP 10-NOV-1999;
 XX 99WO-US26505.
 XX 10-NOV-1998;
 PR 98US-0107840.
 PR 10-MAR-1999;
 PR 99US-0123651.

XX (UTNC-) UNIV NORTH CAROLINA.
 XX PA Rabinowitz JE, Samulski RJ, Xiao W;
 XX PI WPI; 2000-376523/32.
 XX DR Recombinant parvoviral vectors with altered packaging, tropisms and
 XX PT immunogenic properties, useful in gene therapy protocols -
 XX PS Example 27; Page 143-147; 153pp; English.
 XX CC The patent discloses modified parvovirus vectors with advantageous
 CC antigenic properties, packaging capabilities and cellular tropisms.
 CC These vectors can be used in standard recombinant DNA protocols e.g. gene
 CC therapy for delivering nucleic acids to cells. The present sequence is
 CC a helper plasmid encoding a chimeric adeno-associated virus serotype 2/3
 CC (AAV2/3) capsid. The sequence contains the AAV2 rep coding sequences,
 CC most of the AAV2 capsid coding sequences with the exceptions that
 CC loops 2-4 from the AAV2 VP3 subunit is replaced with the corresponding
 CC region from AAV3, in a pBluescript backbone. The rep
 CC coding sequence encodes four proteins responsible for replication,
 CC Rep 68, Rep 78, Rep 52 and Rep 40, and the capsid (cap) coding region
 CC encodes three structural proteins VP1, VP2 and VP3.
 CC Recombinant parvovirus comprising the chimeric capsid is useful
 CC for gene delivery.
 XX SQ Sequence 8178 BP; 2082 A; 2078 C; 2027 G; 1991 T; 0 other;
 Alignment Scores:
 Pred. No.: 1.66e-221 Length: 8178
 Score: 2801.00 Matches: 509
 Percent Similarity: 90.33% Conservative: 33
 Best Local Similarity: 84.83% Mismatches: 56
 Query Match: 86.16% Indels: 2
 DB: 21 Gaps: 2
 US-09-807-802A-15 (1-599) x AAD00834 (1-8178)
 QY 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSer 20
 Db 2544 ACGGCTCCGGGAAAAAAGAGCGCGGTAGACACTCTCTGTGGAGCCAGACTCTCTCTCG 2603
 QY 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLysArgLysPheGlyGlnThrGly 40
 Db 2604 GGAACCGGAAGGCGCGCGCGAGCAGCTGCAAGAAAAAGATTGAATTTGGTCAGACTGA 2663
 QY 41 AspSerGlySerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
 Db 2664 GACGAGACTCAGTACCTGACCCCGAGCTCTCGAGACGCCACCGAGCCCTCTGTGT 2723
 QY 61 ValGlyProThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGluGly 80
 Db 2724 CTGGGAACTAATACGATGGCTACAGGCGAGTGGCGGCCCACTAATGGCAGACATTAACGAGGC 2783
 QY 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
 Db 2784 GCGACGGAGTGGGTAATTCCTCCGAAATTCGCAATTCGATTCACATGGATGGCGGAC 2843
 QY 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyr 120
 Db 2844 AGAGTCATCACCCAGCAGCACCCGAACTGGGGCCCTGCCGCCACTCAACACCAACCACTCTAC 2903
 QY 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer 140
 Db 2904 AACAAATTTCCAGCCATCA---GGAGCTCGACAGACACATCACTACTTTGGCTACAGC 2960
 QY 141 ThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
 Db 2961 ACCCTTGGGGGTATTTTGACTTCAACAGATTCCACTGCGCACTTTTCCACAGCTGACTGG 3020
 QY 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180

Db 3021 CAAGACATCATCAACAACACTGGGGATTCCGACCCAGAGACTCAACTTCAAGCTCTTT 3080
Qy 181 AsnIleGlnValIleGluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeu 200
Db 3081 AACATTCAGTCAAGAGGTCACGCAGAAATGACGGTACGACGATTCGCAATAACCTT 3140
Qy 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
Db 3141 ACCAGCAGGGTTTACTACTCGGAGTACCGAGTCCCGTACGTGCTCGGGTCG 3200
Qy 221 AlaHisGlnGlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGly 240
Db 3201 GCGCACCAAGGCTGCTCCGCGCTTCCAGCGGAGCTTTCATGGTCCCTCAGTATGGA 3260
Qy 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
Db 3261 TACCTCACCTCGAACACGGAAGTCAAGCGGTGGGAGCTCATCTTTTACTGCTGGAG 3320
Qy 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
Db 3321 TACTTCCCTTCGCAGATGCTAAGGACTGGAATAACTTCCAATTCCAGCTATACCTTCGAG 3380
Qy 281 GluValProPheHisSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
Db 3381 GATGTACCTTTTCACAGCAGCTACGCTCACGCCAGAGTTGGATCGCTTGATGATCCT 3440
Qy 301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGln---AsnGlnSerGlySerAla 319
Db 3441 CTTATTGATCAGTATCTGTACTACTACCTGAACAGAACGCAAGGAACAACCTCTGGAACAACC 3500
Qy 320 GlnAsnLysAspLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLys 339
Db 3501 AACCAATCACGGCTGCTTTTACGAGGCTGGGCTCAGTCTATGCTTTTGAGGCCAGA 3560
Qy 340 AsnTrpLeuProGlyProCysTyrArgGlnArgValSerLysThrLysThrAspAsn 359
Db 3561 AATTGGCTACCTGGGCTGCTACCGCCACAGAGACTTCAAAGACTGCTAAACGACAAAC 3620
Qy 360 AsnAsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSer 379
Db 3621 AACCAACAGTAACCTTCTTGGACAGCGCCAGCAAAATATCTCTCAATGGCGGACTCG 3680
Qy 380 IleIleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePro 399
Db 3681 CTGGTGAATCCAGGACCGACTATGCCAGTCAACAAGGACGATGAAGAAATAATTTTCCCT 3740
Qy 400 MetSerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAsp 419
Db 3741 ATGCACGGCAATCTAATATTGGCAAGAGGAGGACACCGCAAGTAACGCAGAAATTAGAT 3800
Qy 420 AsnValMetIleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArg 439
Db 3801 AATGTAATGATTACGAGTGAAGAAGAGATTTCGTACCAACCAATCTCTGTGGCAACAGAGCAG 3860
Qy 440 PheGlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspVal 459
Db 3861 TATGGAACCTGGGCAATAACTTGCAGAGCTCAATATACAGTCCCAAGCTGGAAGTCTGTC 3920
Qy 460 HisAlaMetGlyValAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGly 479
Db 3921 AATCATCAGGGGGCTTACTTGGCATGGTGTGGCAAGATCGTGCAGTGTACCTTCAAGGA 3980
Qy 480 ProIleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGly 499
Db 3981 CCTATCTGGGCAAGATTCTCACAGGATGGACACTTTTCATCTCTCTCTCTGATGGGA 4040
Qy 500 GlyPheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAla 519
Db 4041 GGCCTTGGACTGAACATCCGCTCTCAATCATGATCAAAATACTCCGGTACTCGG 4100
Qy 520 AsnProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThr 539
Db 4101 AATCCTTCGACCACCTTTCAGTGGGCAAGATTGCTTCTTCTTCTTCTCACAGTACTCCAGC 4160

Qy 540 GlyGlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTyrAsn 559
Db 4161 GGCAGGCTCAGCGTGGAGATCGAGTGGAGCTGCAGAGGAAACACGAAACGCTGGAAT 4220
Qy 560 ProGluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAsp 579
Db 4221 CCGAAATTCAGTACACTTCCAACTACACAGTCTGTTAATCGTGAGCTTACCGTGGAT 4280
Qy 580 AsnAsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db 4281 ACTAATGGCTGATTTCAGAGCGCTCGCCCATTTGGCACACAGTACCTGACTCGTAATCTG 4340

RESULT 8

AAH26327 standard; DNA; 4072 BP.

XX AAH26327;

XX 02-OCT-2001 (first entry)

XX Adenovirus helper Ad cap2.

XX AAV; vector; adenovirus; helper virus; Ad cap2; loxP site; gene therapy; ds.

XX Chimeric - Mastadenovirus.

XX Chimeric - Adeno associated virus.

XX Chimeric - Human cytomegalovirus.

XX WO200155361-A2.

XX 02-AUG-2001.

XX 26-JAN-2001; 2001WO-US02709.

XX 26-JAN-2000; 2000US-0178536.

XX (CHIR) CHIRON CORP.

XX Hardy SF;

XX WPI; 2001-483239/52.

XX Producing recombinant adeno-associated virus (rAAV) vector, by stably infecting eukaryotic host cell with rAAV vectors, helper adenovirus or herpes virus, recombinant adenovirus or herpes vectors -

XX Example 2; Page 58-59; 63pp; English.

XX The present sequence is that of adenovirus helper Ad cap2, made by Cre/lox recombination and expressing adeno-associated virus (AAV) VPI, 2.3 from the cytomegalovirus immediate early promoter of pAdlox, and containing a loxP site. Ad cap2 was used to demonstrate recombinant AAV (rAAV) production from virally transduced cells. The invention provides methods and compositions for producing rAAV vector particles by: (a) introducing into a host cell (i) AAV packaging plasmid pFloxAAV (see AAH26326), (ii) a recombinant viral vector encoding plasmid, and (iii) a plasmid encoding herpesvirus, cytomegalovirus or adenoviral functions, or a herpesvirus, cytomegalovirus or adenovirus itself, to produce a flux rAAV particles and rAAV particles; and (b) introducing into a second host cell (i) the rAAV particles or (a), (ii) a vector that directs expression of Cre, and (iii) a vector which directs expression of herpes virus, cytomegalovirus or adenovirus helper functions, such that rAAV vector particles are produced. The vectors are useful for in vivo or in vitro gene therapy and also for in vitro recombinant protein production.

SQ Sequence 4072 BP; 1073 A; 1025 C; 992 G; 982 T; 0 other;

Alignment Scores:

Pred. No.: 6.84e-219 Length: 4072

Score: 2764.50 Matches: 496
Percent Similarity: 89.98% Conservative: 43
Best Local Similarity: 82.80% Mismatches: 59
Query Match: 85.04% Indels: 1
DB: 22 Gaps: 1

US-09-807-802a-15 (1-599) x AAH26327 (1-4072)

QY 1 ThrAlaProGlyValLysValArgProValGluGlnSerProGlnGluProAspSerSerSer 20
DB 1895 ACGGCTCCGGGAAAAAGAGCGGTAGAGCACTCTCTGTGGAGCCAGACTCCCTCCG 1954
QY 21 GlyIleGlyThrGlyGlnGlnProAlaLysLysArgLeuAsnPhelGlyGlnThrGly 40
DB 1955 GGAACGGGAAGCGGGCCAGCAGCTGCAGAAAAAGATTGAATTTGGTCAGACTGGA 2014
QY 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
DB 2015 GACGAGACTCAGTACCTGACCCCGAGCTCTCGGACAGCCACAGCAGCCCTCTGGT 2074
QY 61 ValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnGluGly 80
DB 2075 CTGGGAACCTAATACGATGCTACAGCAGTGGCGCACCAATGGCAGACAATAACAGGGC 2134
QY 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTripleuGlyAsp 100
DB 2135 GCGGACGGAGTGGTAAATCTCTGGGAATTTGGCATTGGATTCCACATGGATGGCGAC 2194
QY 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrThrAsnAsnHisLeuTyr 120
DB 2195 AGAGTCATCACCAACAGCAGCAGCTGGGCGCTGCCACCTACACCAACACCACTCTAC 2254
QY 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer 140
DB 2255 AAACAAATTTCCAGCGCAATCA---GGAGCTCTGAGAGCAATCACTACTTTGGCTACAGC 2311
QY 141 ThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
DB 2312 ACCCTTTGGGGTATTTTGNCTTCAACAGATTTCCACTGCCACTTTTCCACACGTGACTGG 2371
QY 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPhelLysLeuPhe 180
DB 2372 CAAAGACTCATCAACAACTGGGGATTCGACACATTTCCAGTCCACTTTTCAAGCTCTTT 2431
QY 181 AsnIleGlnValLysGluValThrAsnAspGlyValThrIleAlaAsnAsnLeu 200
DB 2432 AACATTCAGTCAAGAGGTTCGCGAATGACGATGACGACGACGATTCGCAATACCTT 2491
QY 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
DB 2492 ACCAGCACGGTTCCAGGTGTTTACTGACTCGAGTACACAGCTCCCGTACCTCTCGGCTCG 2551
QY 221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGly 240
DB 2552 GCGCATCAAGAGTGCCTCCCGCGTTTCCAGCAGACGTCTTTCATGGTGCACAGTATGGA 2611
QY 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
DB 2612 TACCTCACCTGTAACCAACGGAGTACGAGTACGAGTACGAGTCTTCAATTTACTGCTGAG 2671
QY 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
DB 2672 TACTTTCTTCTCAGATGCTGCTACCGGAAACAACTTTTACCTTACCTTACCTTACCTT 2731
QY 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
DB 2732 GAGCTTCCTTTCCACAGCAGCTACGCTCAGCCAGCAGCTCGACCGCTCTCATGAATCCT 2791
QY 301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
DB 2792 CTATCGACAGTACTGCTGATTTACTTTAGCAGCAACAAACACTCCAGTGGACCAACG 2851
QY 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340

DB 2852 CAGTCAAGGCTTCAGTTTCTCAGCGCGGAGCGAGTGCATTTGGGACCACTCTAGGAAC 2911
QY 341 TripleuProGlyProCysTyrArgGlnArgValSerLysThrLysThrAspAsnAsn 360
DB 2912 TGGCTTCTCGACCCCTGTTTACCGCCAGCAGCGAGTATCAAGAGCATCTCGCGATTAACAC 2971
QY 361 AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 380
DB 2972 AACAGTGAATACTCGTGAGCTGGAGCTACCAAGTACCACTCAATGGCAGAGACTCTCTG 3031
QY 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet 400
DB 3032 GTGAATCCGGCCCGCCATGGCAAGCCACAAGAGCATGAAGAAAGTTTTCCTCAG 3091
QY 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
DB 3092 AGCGGGTTCATCTTTTGGGAAGCAAGGCTCAGAGAAAACAAATGTGGACATTTGAAAG 3151
QY 421 ValMetIleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
DB 3152 GTCATGATTACAGACGAGAGGAATCAGGACCAACCAATCCCGTGGCTACGAGCAGTAT 3211
QY 441 GlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHis 460
DB 3212 GGTCTGTATCTACCAACCTCCAGAGAGGCAACAGCAGCAGCTACCGCAGATGTCAAC 3271
QY 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro 480
DB 3272 ACACAAAGCGTTCCTCAGGATGCTCTGGCAGACAGAGATGTATCTTCAGGGGCC 3331
QY 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyLys 500
DB 3332 ATCTGGGCAAGATTCCACACAGCAGCAGCATTTTCCACCCCTCTCCCTCATGGGTGA 3391
QY 501 PheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsn 520
DB 3392 TTCGAGCTTAAACACCTCTCTCCAGATTTCTCATAGAAACACCCCGGTACCTGCGAAT 3451
QY 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
DB 3452 CCTTCGACCACTTCAGTGGCGCAAGTTTCTTCTTCATCACACAGTATCCACGGGA 3511
QY 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPro 560
DB 3512 CAGGTCAGCTGGAGATCGAGTGGAGCTGCAGAGGAAACAGCAACGCTGGAATCCC 3571
QY 561 GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsn 580
DB 3572 GAAATTCAGTACACTTCAACTACCAAGTCTGTAAATGTGGACTTTTACTGTGGACACT 3631
QY 581 AsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
DB 3632 AATGGGTGATTTCAGACCTTCGCCCTCGCCCATTTGGCACCAGATACCTGACTCTGTAATCTG 3688

RESULT 9
AAI66974
ID AAI66974 standard; DNA; 4679 BP.
XX
AC AAI66974;
XX
DT 11-FEB-2002 (first entry)
XX
DE Adeno-associated virus 2 complete genomic sequence.
XX
KW Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytostatic;
KW inverted terminal repeat; nontropic; neuroprotective; antineoplastic; ITR;
KW antidiabetic; antitumor; gene therapy; adeno-associated virus; AAV; db.
XX
OS Adeno-associated virus 2.
XX
FH Location/Qualifiers
FT 321..2252


```

Db 3511 CTCACGACCACTGCTATTTCTTGTAGCAGCAAAACACTCCCAAGTGGAAACCAACCAACG 3570
Qy 321 AenLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAen 340
Db 3571 CAGTCAAGGCTTCACTTTCTCAGCCGCGAGCGAGTGCATTCGGGACGACTAGGAAC 3630
Qy 341 TrpLeuProGlyProCysTyArgGlnGlnArgValSerLysThrLysThrAspAenAen 360
Db 3631 TGGCTTCTTGACCTGTATACCGCCGAGCGAGTATCAAGACATCTCGCGATCAACAC 3690
Qy 361 AenSerAenPheThrTrpThrGlyAlaSerLysTySerAenLeuAenGlyArgGluSerIle 380
Db 3691 AACAGTGAATACTCGTGGACTGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTG 3750
Qy 381 IleAenProGlyThrAlaMetAlaSerHisLysAspAenGluAenLysPhePheProMet 400
Db 3751 GTGAATCCGGGCCCGCCGATGCGCAAGCCACCAAGGACGATGAAGAAAGTTTTTTCCTCAG 3810
Qy 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAenThrAlaLeuAspAen 420
Db 3811 ACGGGGGTCTCATTTTGGAGCAAGGCTCAGAGAAACAAATGTCGACATTTGAAGAAG 3870
Qy 421 ValMetIleThrAspGluGluLysAlaThrAenProValAlaThrGluArgPhe 440
Db 3871 GTCATGATTACAGCAAGAGAAATCAGGACCAACCAATCCGCTGGCTACGAGCAGTAT 3930
Qy 441 GlyThrValAlaValAenPheGlnSerSerSerThrAspProAlaThrGlyAspValHis 460
Db 3931 GGTCTGTATCTACCAACCTCCAGAGAGGCAACAGCAAGCAGCTACCGCAGATGTCAAC 3990
Qy 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTrpLeuGlnGlyPro 480
Db 3991 ACACAAAGGGCTTCTCCAGGCATGCTTGGCAGGACAGAGATGTGTACTTCAGGGGGCC 4050
Qy 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
Db 4051 ATCTGGGCAAGATTCCACACGCGAGCGACATTTTCCACCCCTCTCCCTCATGGGTGA 4110
Qy 501 PheGlyLeuLysAenProProGlnIleLeuLysAenThrProValProAlaAen 520
Db 4111 TTCGGAATTAACACCCCTCTCCACAGATTCTCATCAAGAACACCCCGGTACTTCGGAAT 4170
Qy 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTySerThrGly 540
Db 4171 CTTTCGACCACTTCAGTGGCGCAAGTTTGTCTTCTTCATCACACAGTACTCCACGGGA 4230
Qy 541 GlnValSerValGluLysGluTrpGluLeuGlnLysGluAenSerLysArgTrpAenPro 560
Db 4231 CAGGTCAAGCTGAGATCAGTGGGAGCTGCAGAGAGGAAACAGCAAAACGCTGGAAATCCC 4290
Qy 561 GluValGlnTyThrSerAenTyAlaLysSerAlaAenValAenValAenValAenValAen 580
Db 4291 GAAATTCAGTACACTTCCAACTACAAAGTCTGTATATGTGGACTTTACTGTGGACACT 4350
Qy 581 AenGlyLeuTyThrGluProArgProIleGlyThrArgTyLeuThrArgProLeu 599
Db 4351 AATGGGCTGTATTCAGAGCTTCGCCCATTTGGCACCAGATACCTGACTCGTAATCTG 4407

RESULT 10
AAF23750
ID AAF23750 standard; DNA; 4679 BP.
XX
AC AAF23750;
XX
DT 28-MAR-2001 (first entry)
XX
DE AAV2 DNA sequence.
XX
KW AAV2; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
KW atherosclerosis; sickle cell anemia; thalassemia;
KW blood clotting disorder; diabetes; ss.
XX

```

```

OS Adeno associated virus.
XX US6156303-A.
XX
XX PD 05-DEC-2000.
XX
XX PF 11-JUN-1997; 97US-0873168.
XX
XX PR 11-JUN-1997; 97US-0873168.
XX
XX PA (UNIW ) UNIV WASHINGTON.
XX
XX PI Russell DW, Rutledge EA;
XX WPI; 2001-060164/07.
XX
XX DR
XX
XX PT Adeno-associated virus serotype 6 and viral vector derived from it for
XX gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
XX syndrome, sickle cell anemia, thalassemia and diabetes -
XX
XX PS Claim 7; Fig 1; 50pp; English.
XX
XX CC The present invention relates to adeno-associated virus serotypes. The
XX present sequence is the DNA sequence of one such serotype (AAV2). AAV2
XX can be used to construct AAV viral vectors for use in gene therapy for a
XX range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
XX sickle cell anaemia, thalassemia, blood clotting disorders and diabetes.
XX The AAV viral vectors have increased transduction efficiency of a
XX particular host cell as the AAV virion containing the AAV vector genome
XX can be modified to express a capsid protein of an AAV serotype that
XX transduces the selected host cell.
XX
XX SQ Sequence 4679 BP; 1197 A; 1263 C; 1255 G; 964 T; 0 other;

Alignment Scores:
Pred. No.: 8.26e-219 Length: 4679
Score: 2764.50 Matches: 496
Percent Similarity: 89.98% Conservative: 43
Best Local Similarity: 82.80% Mismatches: 59
Query Match: 85.04% Indels: 1
DB: 22 Gaps: 1

US-09-807-802A-15 (1-599) x AAF23750 (1-4679)
Qy 1 ThrAlaProGlyLysLysArgProValGlnSerProGlnGluProAspSerSerSer 20
Db 2614 ACAGCTCCGGGAAAAAGAGGCGGTAGAGCACTCTCTGTGGAGCCAGACTCTCTCTCG 2673
Qy 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAenPheGlyGlnThrGly 40
Db 2674 GGAACCGGAAAGCGGGCCAGCAGCTGCAAGAAAAAGATTGAATTTTGGTCACATGGA 2733
Qy 41 AspSerGluSerValProAspProGlnProLeuGlyGluProAlaThrProAlaAla 60
Db 2734 GACGACGACTCAGTACCTTGACCCCGCTCTCGACAGCCACCCAGCAGCCCTCTCTGT 2793
Qy 61 ValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAenAenGluGly 80
Db 2794 CTGGGAATAATACGATGGCTACAGGCACTGGCGCAATGGCAGCAATAAAGAGGGC 2853
Qy 81 AlaAspGlyValGlyAenAlaSerGlyAenTrpHisCysAspSerThrTrpLeuGlyAen 100
Db 2854 GCCGACGGAGTGGGTAAATTCCTCCGAAATTTGGCATTTCCGATTCACATGATGGGGC 2913
Qy 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTyAsnAenHisLeuTy 120
Db 2914 AGAGTCATCACCAACGACGCCGAACTTGGGCCCTTCCCACTACCAACCACTCTTAC 2973
Qy 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAenAenAenHisTyPheGlyTySer 140
Db 2974 AAACAAATTTCCAGCCAAATCA---GGAGCTTCGAACGACAAATCACTACTTTGGCTACAGC 3030
Qy 141 ThrProTrpGlyTyThrPheAspPheAsnArgPheHisCysHisPheSerProArgSerTrp 160

```

```
Db 3031 ACCCCCTTGGGGGATTTTGGCTTCAACAGATTCCACTGCCACTTTTACACACGCTGACTGG 3090
Qy 161 GlnArgLeuIleAsnAsnAsnTyrGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
Db 3091 CAAGACTCATCAACAACCACTGGGGATTCCGACCCCAAGAGACTCAACTTCAAGCTCTTT 3150
Qy 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeu 200
Db 3151 AACATTCAAGTCAAAAGAGTCCAGCAAGTACCGGTACGACGATTCGCAATTAACCTT 3210
Qy 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
Db 3211 ACCACACGGTTCAGGTGTTTACTGACTCGGAGTACCAGCTCCCGTACTGCTCTGGCTCG 3270
Qy 221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGly 240
Db 3271 GGGCATCAAGGATGCTCCCGCGCTTCCAGCAGACGCTTTCATGGTGCCACAGTATGGA 3330
Qy 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
Db 3331 TACCTCACCTCGAACACAGCGGAGTCAGGACGATGAGACGCTCTTCAATTTTACTGCTGGAG 3390
Qy 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
Db 3391 TACTTTCCTTTCAGATGCTGGTACCGGAACAACATTTTACTTTCAGCTACACTTTTGG 3450
Qy 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
Db 3451 GACGTTCTTTCACAGCAGCTACGCTCACAGCCAGAGTCTGGACCGCTCATGAATCCT 3510
Qy 301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
Db 3511 CTCATCGACCACTGCTGTTTACTGTGAGCAGAACAAACACTCCAAAGTGGAAACACCCACG 3570
Qy 321 AsnLysAspLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
Db 3571 CAGTCAAGCTTCAGTTTCTCAGCCCGGAGCGAGTGACATTCGGGACCATCTAGGAAAC 3630
Qy 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
Db 3631 TGGCTTCTGGACCTGTTTACCGCAGCAGCGAGTATCAAGACATCTCGCGGATAACAAC 3690
Qy 361 AsnSerAsnPheThrTyrThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 380
Db 3691 AACAGTGAATATCTGTGACTGGAGCTACCAAGTACCACTCAATGGCAGAGACTCTCTG 3750
Qy 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet 400
Db 3751 GTGATCCGGCCCGCCCATGGCAAGCCACCAAGGACGATGAAGAAAAGTTTTTCTCTCAG 3810
Qy 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
Db 3811 AGCGGGTCTCATCTTTGGGAAGCAAGGCTCAGAGAAAACAAATGTGGACATTCAAAAG 3870
Qy 421 ValMetIleThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
Db 3871 GTCATGATTACAGACGAAGAGAAATCAGGACCAACCAATCCCGTGGCTACCGAGCAGTAT 3930
Qy 441 GlyThrValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHis 460
Db 3931 GGTTCGTGATCTACCACTCCAGAGAGGCAACAGACGAGCTACCGCAGATGTCAAC 3990
Qy 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro 480
Db 3991 ACACAAGGCTTCTTCCAGGATGCTCTGGCAGGACAGAGATGTGTACTTTCAGGGGCC 4050
Qy 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetCysGly 500
Db 4051 ATCTGGCAAGAGATTCCACACAGCAGGACGACATTTTCACTCCCTCTCCCTCATGGTGA 4110
Qy 501 PheGlyLeuLysAsnProProGlnIleLeuLysAsnThrProValProAlaAsn 520
```

```
Db 4111 TTCGACTTAACACACCTCTCTCCAGATTCTCATCAAGAACACCCCGGTACCTGGCAAT 4170
Qy 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
Db 4171 CCTTCGACCACTTTCAGTGGCGGCAAGTTTGTCTTCTTCATCACACAGTACTCCACGGA 4230
Qy 541 GluValSerValGluIleGluTyrGluLeuGlnLysGluAsnSerLysArgTrpAsnPro 560
Db 4231 CAGGTCAGCGTGGAGATCGAGTGGGAGCTGCAGAGGAAACAGCAACCGCTGGAATCCC 4290
Qy 561 GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsn 580
Db 4291 GAAATTCAGTACACTTCCNACTACACAAGTCTGTTAATGTGGACTTTTACTGTGGACACT 4350
Qy 581 AsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db 4351 AATGGCGTGATTTCAGAGCCTCGCCCATTTGGCACCAGATACCTGACTCGTAATCTG 4407
RESULT 11
ABK89694
ID ABK89694 standard; DNA; 4679 BP.
AC ABK89694;
XX
DT 05-NOV-2002 (first entry)
XX
DE Adeno-associated virus 2 (AAV2) vector.
XX
KW Adeno-associated virus 2 vector; AAV2; ds; cyclic; circular; cancer;
KW VP1 capsid; heparin-sulphate proteoglycan; vaccine; immune response;
XX ovarian cancer.
XX
OS Adeno-associated virus 2.
XX
FH Key Location/Qualifiers
FT CDS 2203..4410
FT /tag= a
FT /product= "Adeno-associated virus 2, VP1 capsid
FT protein"
FT CDS 2614..4410
FT /tag= b
FT /product= "Adeno-associated virus 2, VP2 capsid
FT protein"
FT CDS 2809..4410
FT /tag= c
FT /product= "Adeno-associated virus 2, VP3 capsid
FT protein"
XX
PN WO200253703-A2.
XX
PD 11-JUL-2002.
XX
PF 04-JAN-2002; 2002WO-US00152.
XX
PR 05-JAN-2001; 2001US-260124P.
XX
PA (CHIL-) CHILDRENS HOSPITAL INC.
XX
PI Bartlett JS;
XX
DR WPI; 2002-583608/62.
DR P-PSDB; AAU98974, AAU98975, AAU98976.
XX
PT New adeno-associated virus vector comprises a biotinylated capsid or
PT capsid protein with an amino acid insertion in the VP1 capsid, useful
PT as a vaccine or for transferring a therapeutic peptide to a cancer cell
PT
XX
PS Disclosure; Page 46-48; 57pp; English.
XX
CC The invention relates to an adeno-associated virus (AAV) vector (I)
comprising a biotinylated capsid or capsid protein (II) with an amino
```

CC acid insertion following the capsid amino acid at position 139, 161, 588
 CC or 657 in the VP1 capsid. The AAV vector comprises a capsid protein
 CC containing one or more amino acid insertions that ablate the ability of
 CC the vector to bind heparin-sulphate proteoglycan and allow the ability to
 CC use a cellular receptor not used by wild type AAV. Modified (I) are
 CC useful as vaccines to elicit immune responses to amino acids, where the
 CC response can be protective and/or therapeutic. (II) may be used to
 CC transfer a therapeutic peptide to a cancer cell, particularly to an
 CC ovarian cancer cell. The present sequence represents the adeno-associated
 CC virus 2 (AAV2) vector used in to make modified AAV2 vectors.

XX Sequence 4679 BP; 1198 A; 1262 C; 1255 G; 964 T; 0 other;

Alignment Scores: Length: 4679
 Pred. No.: 8,266-219 Matches: 496
 Score: 2764.50
 Percent Similarity: 89.98% Conservative: 43
 Best Local Similarity: 82.80% Mismatches: 59
 Query Match: 85.04% Indels: 1
 DB: 24 Gaps: 1

US-09-807-802a-15 (1-599) x ABK89694 (1-4679)

QY 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSer 20
 DB 2614 ACGGCTCCGGGAAAAAGAGCGCGTAGAGCACTCTCTGTGGAGCGAGACTCTCTCTCG 2673
 QY 21 GlyIleGlyThrGlyGlnGlnProAlaLysLysArgLysAspPheGlyGlnThrGly 40
 DB 2674 GGAAACCGGAAGCGGGCGAGAGCTGCAAGAAAAAGATTGAATTTTGGTCAGACTGGA 2733
 QY 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
 DB 2734 GAGCGAGACTAGTACTGACCCCGAGCTCTCGGACAGCCAGCCAGCCCTCTGCT 2793
 QY 61 ValGlyProThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAspAsnGluGly 80
 DB 2794 CTGGGAACCTAATACGATGGCTACAGGCGAGTGGCGCACCAATGGCAGACAATAACAGGCG 2853
 QY 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
 DB 2854 GCGGAGCGAGTGGTAAATCTCTCGGAAATTTGGCATTTGCATTTCCATGGATGGCGGAC 2913
 QY 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTrpAsnAsnHisLeuTyr 120
 DB 2914 AGAGTCATCACCAAGCAGCCGGAACCTGGGCGCTCCACCTTACAAACACCACTCTAC 2973
 QY 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer 140
 DB 2974 AAACAAATTTCCAGCCCAATCA---GGAGCGCTCGAAGCAGCAATCACTACTTTGGCTACAGC 3030
 QY 141 ThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTyr 160
 DB 3031 ACCCTTGGGGGATTTTGATCTTCAACAGATTCCATCGCACCTTTTCCACACGAGTGG 3090
 QY 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
 DB 3091 CAAAGACTCATCAACAACTGGGATTCGGACCCCAAGAGACTCAACTTCAAGCTCTTT 3150
 QY 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeu 200
 DB 3151 AACATTCNAAGTCAAGAGGTTCAGCGAGAATGACGGTACGAGCAGATTCGCCAATAACCTT 3210
 QY 201 ThrSerThrValGlnValPheAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
 DB 3211 ACCAGACGGTTTCAAGTGTATTACTGACTCGAGATCACAGCTCCCGTCTCGCTCGCTCG 3270
 QY 221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGly 240
 DB 3271 GCGCATCAAGGATGCTCCCGCTTCCAGCAGAGAGCTCTTTCATGTGTGCGCACAGTATGGA 3330
 QY 241 TyrIleuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260

DB 3331 TACCTCACTGAAACAAACGCGAGTCAGGCACTAGGACGCTCTCTCATTTTACTGCTGGAG 3390
 QY TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
 DB 3391 TACTTTCCTTCTCAGATGCTCGTACCGGAACAACTTTTACTCTTCAGTACACTTTTGG 3450
 QY GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
 DB 3451 GACGTTCTTCTCCACAGCAGCTACGCTCACAGCCAGAGTCTGGACCGTCTCATGAATCCT 3510
 QY 301 LeuIleAspGlnTyrLeuTyrTyrIleuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
 DB 3511 CTCTACGACCACTACCTGTATTACTTGGAGCAACAAACACTCCCAAGTGGAAACACCCACG 3570
 QY 321 AsnLysAspLeuLeuPheSerArgLysSerProAlaGlyMetSerValGlnProLysAsn 340
 DB 3571 CAGTCAAGGCTTCAGTTTCTCAGCGCGGAGCGAGTGACATTCGGGACCACTCTAGGAAC 3630
 QY 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
 DB 3631 TGGCTTCTCGACCTGTTCACCGCAGCAGCGAGTATCAAAGACATCTGCGGATAACAAC 3690
 QY 361 AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 380
 DB 3691 AACAGTGAATACTCGTGACTGGAGCTACCAAGTACCACTCAATGGCAGAGACTCTCTG 3750
 QY 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet 400
 DB 3751 GTGAATCCGGCGCGCCATGCGCAAGCCACCAAGACGATGAAGAAAAGTCTTTTCTCAG 3810
 QY 401 SerGlyValMetIlePheGlyGlySerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
 DB 3811 AGCGGGTCTCATCTTTGGGAAGCAAGGCTCAGAGAAAACAAATGTGGACATTTGAAAG 3870
 QY 421 ValMetIleThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
 DB 3871 GTCATGATTACAGAGAAAGAGAAATCAGGACAACCAATCCCGTGGCTACGGAGCAGTAT 3930
 QY 441 GlyThrValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHis 460
 DB 3931 GGTTCGTATCTACCACTCCAGAGAGCAACAGACAGCAGCTACCGCAGATGTCAAC 3990
 QY 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrIleuGlnGlyPro 480
 DB 3991 ACACAAGCGTCTTCTCCAGGATGCTGTGGCAGGACAGAGATGTGTACTCTCAGGGGCC 4050
 QY 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
 DB 4051 ATCTGGGCAAGATTCCACACAGCAGCAGACATTTTCACTCCCTCTCCCTCATGGGTGA 4110
 QY 501 PheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsn 520
 DB 4111 TTCGGACTTAACACCTCTCCACAGATTCTCATCAAGAACACACCCCGTACTCGAAT 4170
 QY 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
 DB 4171 CCTTCGACCACTCTCAGTGGCGCAAGTTTGTCTCTTCATCACAGTACTCCACGGA 4230
 QY 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPro 560
 DB 4231 CAGGTCAGCTGGAGATCGATGGGAGCTCGAGAGGAAACAGCAACGCTGGATCCC 4290
 QY 561 GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsn 580
 DB 4291 GAAATTCAGTACACTTCCAACTTACAAAGTCTGTTAATGTGGACTTTTACTCTGGACACT 4350
 QY 581 AsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
 DB 4351 AATGGCGTGTATTACAGGCGCTCGCCCCATTGGCAGCATACCTGACTCGTAATCTG 4407

RESULT 12

ID ABV76133 standard; DNA; 4679 BP.

QY 421 ValMetIleThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
 DB 3871 GTCATGATTACAGACGAGGAGAAATCAGCAACCAATCCGCTGGCTACGGAGCAGTAT 3930
 QY 441 GlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHis 460
 DB 3931 GGTTCGTATCTACCAACCTCCAGAGAGGCAACAGACAGCTACCCGACAGATGTCAAC 3990
 QY 461 AlaMetGlyValAlaLeuProGlyMetValTrpGlnAspArgAspValTrpLeuGlnGlyPro 480
 DB 3991 ACACAGAGGGCTTCTCCAGGCATGCTGCGCAGGACAGAGATGTGTACCTTCAGGGGCC 4050
 QY 481 IleTrpAlaValIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
 DB 4051 ATCTGGGCAAGATTCCACACGCGACGACATTTTTCACCCCTCTCCCTCATGGGTGA 4110
 QY 501 PheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsn 520
 DB 4111 TTGGGACTTAAACACCTCTCTCCACAGATTCTCATCAAGACACACCCCGTACTCGGAAT 4170
 QY 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTrpSerThrGly 540
 DB 4171 CCTTCGACCACTTCAGTCGGCAAGTTTGTCTTCTTCATCACACAGTACTCCACGGGA 4230
 QY 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPro 560
 DB 4231 CAGTCCAGCGTGAGATCAGTGGGAGCTGCAGAGCAAAACAGCAACCTGGGAATCCC 4290
 QY 561 GluValGlnTrpThrSerAsnTrpAlaLysSerAlaAsnValAspPheThrValAspAsn 580
 DB 4291 GAAATTCAGTACTTCCAACTACAAAGTCTGTTAATGTGGACTTTACTGTGGACACT 4350
 QY 581 AsnGlyLeuTrpThrGluProArgProIleGlyThrArgTrpLeuThrArgProLeu 599
 DB 4351 AATGGCGTGATTTCAGAGCCTCGCCCAATGGCACCAGATACCTGACTCGTAATCTG 4407
 RESULT 13
 ID AAH26326
 XX AAH26326 standard; DNA; 7557 BP.
 AC AAH26326;
 XX
 DT 02-OCT-2001 (first entry)
 XX
 XX Adeno-associated virus packaging plasmid pfloxAAV.
 DE
 XX AAV; vector; pfloxAAV2; packaging cell line; gene therapy; ds.
 XX
 XX Adeno associated virus.
 OS
 XX WO200155361-A2.
 PN
 XX 02-AUG-2001.
 PD
 XX
 XX 26-JAN-2001; 2001WO-US02709.
 XX
 XX 26-JAN-2000; 2000US-0178536.
 XX
 XX (CHIR) CHIRON CORP.
 PA
 XX Hardy SF;
 XX
 XX WPI; 2001-483239/52.
 DR
 XX Producing recombinant adeno-associated virus (rAAV) vector, by stably
 PT infecting eukaryotic host cell with rAAV vectors, helper adenovirus or
 PT herpes virus, recombinant adenovirus or herpes vectors -
 XX
 XX Example 1; Page 54-57; 63pp; English.
 PS
 XX The present sequence is that of adeno-associated virus (AAV)
 XX packaging plasmid pfloxAAV2. The following AAV packaging plasmids
 CC

CC were prepared: pKsrepcap, a non-replicating control plasmid
 CC containing rep and cap genes but no inverted terminal repeats
 CC (ITRs); pAV2 (see AAH26325), containing a wild-type AAV genome; and
 CC pfloxAAV, a version of pAV2 with 2 loxP sites inserted such they
 CC flank the rep and cap genes and thus separate the ITRs from these
 CC genes. Each was combined with pCMV GFP (see AAH26324), a plasmid
 CC with a green fluorescent protein expressing recombinant AAV (rAAV)
 CC genome, and used to transfect 293 or Cre8 cells. After 6 hr, the
 CC medium was changed and wild-type adenovirus type 5 was added.
 CC virus particles were harvested 3 days later. Plasmid pAV2
 CC produced mostly AAV and a low yield of GFP vector with no effect
 CC on Cre recombinase. pfloxAAV packaged as much GFP vector as the
 CC control plasmid and while the amount of GFP vector was not
 CC affected by Cre recombinase, the amount of floxAAV in the product
 CC was reduced to 1/10 of the 293 value by the action of Cre
 CC recombinase. In 293 cells it was evident that the GFP vector did
 CC not compete with wild-type AAV for replication or packaging, while
 CC the GFP vector was able to complete against floxAAV. Methods
 CC and compositions are provided for producing rAAV vector particles
 CC by: (a) introducing into a host cell (i) pfloxAAV, (ii) a
 CC recombinant viral vector encoding plasmid, and (iii) a plasmid
 CC encoding herpesvirus, cytomegalovirus or adenoviral functions, or
 CC a herpesvirus, cytomegalovirus or adenovirus itself, to produce
 CC a flox AAV particles and rAAV particles; and (b) introducing into a
 CC second host cell (i) the rAAV particles or (a), (ii) a vector
 CC that directs expression of Cre, and (iii) a vector which directs
 CC expression of herpes virus, cytomegalovirus or adenovirus helper
 CC functions, such that rAAV vector particles are produced. The
 CC vectors are useful for in vivo or in vitro gene therapy and also
 CC for in vitro recombinant protein production.
 XX

SQ Sequence 7557 BP; 1970 A; 1940 C; 1953 G; 1694 T; 0 other;
 Alignment Scores:
 Pred. No.: Length: 7557
 Score: 2764.50 Matches: 496
 Percent Similarity: 89.98% Conservative: 43
 Best Local Similarity: 82.80% Mismatches: 59
 Query Match: 85.04% Indels: 1
 DB: Gaps: 1

US-09-807-802A-15 (1-599) x AAH26326 (1-7557)
 QY 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSer 20
 DB 2642 ACGGCTCCGGGAAAAAGAGAGCGCGTAGAGCACTCTCTGTGGAGCCAGACTCTCTCTCG 2701
 QY 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGly 40
 DB 2702 GGAACCGGAAAGGCGGCGCAGCAGCTGCAGAAAAAGATTGAATTTTGGTCAGACTGGA 2761
 QY 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
 DB 2762 GACGAGACTCAGTACTTACCTGACCCCGAGCTCTCGACAGCCACAGAGCCCTCTGCT 2821
 QY 61 ValGlyProThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAspAsnGluGly 80
 DB 2822 CTGGGAACCTAATACGATGCTACGAGCAGTGGCGCAACCAATGCGCAGACAAATACGAGGCG 2881
 QY 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
 DB 2882 GCCGACGAGTGGGTAATTCCTCGGGAAATTCGCAATTCGATTCACATGGATGGCGCAG 2941
 QY 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTrpAsnAsnHisLeuTrp 120
 DB 2942 AGAGTCATCAACCCAGCAGCACCAGACCTGGGGCCCTGCCACCTACACCAACCACTCTAC 3001
 QY 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTrpPheGlyTrpSer 140
 DB 3002 AAACAAATTTCCAGCCAAATCA---GGAGCCTCGAAGCAATCACTACTCTTGGCTACAGC 3058
 QY 141 ThrProTrpGlyTrpPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160

| | | | |
|------|----|---|------|
| 3059 | Db | ACCCCTTGGGGGTATTTTGGACTTCAACAGATTCACCTGCGCACCTTTTCCACACGTGACTGG | 3111 |
| 161 | Qy | GlnArgLeuIleAenAenAenTrpGlyPheArgProLysArgLeuAenPheLysLeuPhe | 180 |
| 3119 | Db | CAAGACTCATCAACAACAACCTGGGGATTCGACCCCAAGAGACTCAACTTCAAGCTCTTT | 3178 |
| 181 | Qy | AsnIleGlnValLysGluValThrThrAsnAspGlyValThrIleAlaAenAenLeu | 200 |
| 3179 | Db | AACATTCAAGTCAAGAGAGGTCAACGCAAGATGACGGTACGACGACGATGCGCAATAACCTT | 3238 |
| 201 | Qy | ThrSerThrValGlnValPheSerAspSerGlnTyrGlnLeuProTyrValLeuGlySer | 220 |
| 3239 | Db | ACCAAGCAGGTTCAGGTGTTTACTGACTCGGAGTACCAGCTCCCGTACGTCTCGGCTCG | 3298 |
| 221 | Qy | AlaHisGlnGlyCysLeuProPheProPheProAlaAspValPheMetIleProGlnTyrGly | 240 |
| 3299 | Db | GGCATCAAGGATGCCTCCCGCGTTCCAGCAGACGCTCTTCATGTTACTGCCTGGAG | 3358 |
| 241 | Qy | TyrLeuThrLeuAenAenGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu | 260 |
| 3359 | Db | TACCTCACCTTGAACAACGGGAGTCAGGCGAGTAGGACGCTCTTCATTTTACTGCCTGGAG | 3418 |
| 261 | Qy | TyrPheProSerGlnMetLeuArgThrGlyAsnAenPheThrPheSerTyrThrPheGlu | 280 |
| 3419 | Db | TACTTTCTCTCAGATGCTCGTACCGGAAACAACCTTACCTTCAGCTACACTTTTGAG | 3478 |
| 281 | Qy | GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro | 300 |
| 3479 | Db | GACGTTCTCTTCCACAGCAGCTACGCTCAGCCAGAGTCTGGACCGTCTCATGAATCCT | 3538 |
| 301 | Qy | LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAenGlnSerGlySerAlaGln | 320 |
| 3539 | Db | CTCATCGACCAAGTACCTGTATTACTTCAGCAGAAACAACACTCTCAAGTGAACCAACCAGC | 3598 |
| 321 | Qy | AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn | 340 |
| 3599 | Db | CAGTCAAGGCTTCAGTTTCTTCAGCCGGAGCGAGTGACATCTCGGACCAAGTCTAGGAAC | 3658 |
| 341 | Qy | TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn | 360 |
| 3659 | Db | TGGCTTCTTGACCCCTGTATCCGCGCAGCAGCGAGTATCAAGACATCTCGGATTAACAC | 3718 |
| 361 | Qy | AsnSerAenPheThrTrpThrGlyAlaSerLysTyrAenLeuAenGlyArgGluSerIle | 380 |
| 3719 | Db | AACAGTGAATACTCGTGGACTGGAGCTACCAAGTACCACTCAATGCGAGAGACTCTCTG | 3778 |
| 381 | Qy | IleAenProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet | 400 |
| 3779 | Db | GTGAATCGGGCCCGGCCATCGCAAGCCACAAAGACCATGAAGAAAAGTTTTTTCCTCAG | 3838 |
| 401 | Qy | SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAenThrAlaLeuAspAsn | 420 |
| 3839 | Db | AGCGGGTTCATCTTTTGGAGAGCAAGGCTCAGAGAAAACAATGTGGACATTTGAAAG | 3898 |
| 421 | Qy | ValMetIleThrAspGluGluGluIleLysAlaThrAenProValAlaThrGluArgPhe | 440 |
| 3899 | Db | GTCAATGATTACAGACGAAGAGAAATCAGGACAACCAATCCGTGGCTACGAGCAGTAT | 3958 |
| 441 | Qy | GlyThrValAlaValAenPheGlnSerSerSerThrAspProAlaThrGlyAspValHis | 460 |
| 3959 | Db | GGTTCTGTATCTACCAACCTCCAGAGAGGCACAGACAGCAGCTCCGCGAGATGTCAAC | 4018 |
| 461 | Qy | AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro | 480 |
| 4019 | Db | ACACAGGCGTTCCTCCAGGCATGGTCTGGCAGGACAGAGATGTACCTTCAGGGGCC | 4078 |
| 481 | Qy | IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly | 500 |
| 4079 | Db | ATCTGGGCAAGATTCACACACAGCAGCATTTTACCCCTCTCCCTCATGGGTGGA | 4138 |
| 501 | Qy | PheGlyLeuLysAenProProGlnIleLeuLysAenThrProValProAlaAsn | 520 |
| 4139 | Db | TTCCGACTTAAACACCTCTTCCACAGATCTCATCAAGAACACCCGGTACCTGCGCAAT | 4198 |

CC recombinant viral vector encoding plasmid, and (iii) a plasmid
 CC encoding herpesvirus, cytomegalovirus or adenoviral functions, or
 CC a herpesvirus, cytomegalovirus or adenovirus itself, to produce a
 CC flox AAV particles and rAAV particles; and (b) introducing into a
 CC second host cell (i) the rAAV particles or (a), (ii) a vector
 CC that directs expression of Cre, and (iii) a vector which directs
 CC expression of herpes virus, cytomegalovirus or adenovirus helper
 CC functions, such that rAAV vector particles are produced. The
 CC vectors are useful for in vivo or in vitro gene therapy and also
 CC for in vitro recombinant protein production.

XX SQ Sequence 8698 BP; 2117 A; 2368 C; 2306 G; 1907 T; 0 other;

| Alignment Scores: | | 1,918-218 | Length: | 8698 |
|------------------------|--|-----------|---------------|------|
| Pred. No.: | | 2764.50 | Matches: | 496 |
| Percent Similarity: | | 89.98% | Conservative: | 43 |
| Best Local Similarity: | | 82.80% | Mismatches: | 59 |
| Query Match: | | 85.04% | Indels: | 1 |
| DB: | | 22 | Gaps: | 1 |

US-09-807-802A-15 (1-599) x AAH26325 (1-8698)

| | | | |
|----|------|--|------|
| QY | 1 | ThrAlaProGlyLysArgProValGluGlnSerProGlnGluProAspSerSerSer | 20 |
| DB | 2614 | ACGGCTCCGGGAAAAAAGAGCGGTAGAGCACTCTCTGTGGAGCCAGACTCTCTCTCG | 2673 |
| QY | 21 | GlyIleGlyThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGly | 40 |
| DB | 2674 | GAACCGGAAAGCGCGCCAGCAGCTGCAAGAAAGATTGAATTTGGTCAGACTGGA | 2733 |
| QY | 41 | AspSerGluSerValProAspProGlnProLeuGluGluProAlaThrProAlaAla | 60 |
| DB | 2734 | GACGAGACTAGTACTGACCTCCCGCCAGCTCTCGACAGCCAGCAGCCCTCTGCT | 2793 |
| QY | 61 | ValGlyProThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAspAsnGluGly | 80 |
| DB | 2794 | CTGGGAACATAACGATGCTACAGCGAGTGGCGCACCAATGGCAGACAATAACGAGGC | 2853 |
| QY | 81 | AlaAspGlyValGlyAsnAlaSerGlyAsnThrHisCysAspSerThrTrpLeuGlyAsp | 100 |
| DB | 2854 | GCGGACGAGTGGGTAAATTTCTCGGAAATTTGGCAATTCGATTCACATGGATGGCGAC | 2913 |
| QY | 101 | ArgValIleThrSerThrArgThrTrpAlaLeuProThrTyrrAsnAsnHisLeuTyrr | 120 |
| DB | 2914 | AGAGTCATCACCCAGCACCAGACCTGGGCCCTGCCACCTACACACACCACTCTAC | 2973 |
| QY | 121 | LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrrPheGlyTyrrSer | 140 |
| DB | 2974 | AAACAAATTTCCAGCCCAATCA---GGAGCCTCGAACGACAATCACTACTTTGGCTACAGC | 3030 |
| QY | 141 | ThrProTrpGlyTyrrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp | 160 |
| DB | 3031 | ACCCCTTGGGGGTATTTTGAATTCACACGATTCCTGCTGCTTTCACACGCTGACTGG | 3090 |
| QY | 161 | GlnArgLeuIleAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe | 180 |
| DB | 3091 | CAAGACTCATCAACCAACTGGGATTCGACCCCAAGAGACTCAACTTCAAGCTCTTT | 3150 |
| QY | 181 | AsnIleGlnValLysGluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeu | 200 |
| DB | 3151 | AACATTTCAAGTCAAGAGGTACGCAAGATGACGGTACGACGACGATTCGCAATAACCTT | 3210 |
| QY | 201 | ThrSerThrValGlnValPheSerAspSerGluTyrrGlnLeuProTyrrValLeuGlySer | 220 |
| DB | 3211 | ACCAGCAGGCTTCAAGTGTTTACTGACTCGGAGTACAGCTCCCGTACCTCTCGGCTCG | 3270 |
| QY | 221 | AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrrGly | 240 |
| DB | 3271 | GCGCATCAAGATGCTCCCGCGTTCGCCAGCAGACGCTCTTCATGGTGGCCACAGTATGA | 3330 |
| QY | 241 | TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrrCysLeuGlu | 260 |

RESULT 15

AAAT09008

ID AAAT09008 standard; DNA; 4680 BP.

| | | | |
|----|------|--|------|
| DB | 3331 | TACCTACCCCTGAACAACAGCGAGTCAGCGAGTAGGACGCTCTTTCATTTTACTGCTCGAG | 3390 |
| QY | 261 | TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrrThrPheGlu | 280 |
| DB | 3391 | TACTTTCTCTCTCAGATGCTGCGTACCGGAACAACTTTTACCTTCAGCTACACTTTTGA | 3450 |
| QY | 281 | GluValProPheHisSerSerTyrrAlaHisSerGlnSerLeuAspArgLysLeuMetAsnPro | 300 |
| DB | 3451 | GACGCTCTCTTCCACAGCAGCTAGCTCACAGCCAGAGTCTGGACCGCTCTCATGAATCCT | 3510 |
| QY | 301 | LeuIleAspGlnTyrrLeuTyrrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln | 320 |
| DB | 3511 | CTCATCGACCACTGCTGTATTACTTTCAGCAGAAACAAACACTTCCAAAGTGGAAACCA | 3570 |
| QY | 321 | AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn | 340 |
| DB | 3571 | CAGTCAAGGCTTCAGTTTTTCTCAGCCGCGAGCGAGTACATTCGGGACCACTCTAGGAAC | 3630 |
| QY | 341 | TrpLeuProGlyProCysTyrrArgGlnArgValSerLysThrLysThrAspAsnAsn | 360 |
| DB | 3631 | TGGCTTCTCGACCTGTACCGCCAGCAGCGAGTATCAAGACATCTCGCGATAACAAC | 3690 |
| QY | 361 | AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrrAsnLeuAsnGlyArgGluSerIle | 380 |
| DB | 3691 | AACAGTGAATACTCGTGGACTTGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTG | 3750 |
| QY | 381 | IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet | 400 |
| DB | 3751 | GTGAATTCGGGCCCGCCATGGCAAGCCACNAGACCATGNAAGAAAGTTTTTCTCTCAG | 3810 |
| QY | 401 | SerGlyValMetIlePheGlyLysSerAlaGlyAlaSerAsnThrAlaLeuAspAsn | 420 |
| DB | 3811 | AGCGGGTTCATCTCTTGGGAGCAAGGCTCAGAGAAAACAAATGTGGACATTTGAAAG | 3870 |
| QY | 421 | ValMetIleThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe | 440 |
| DB | 3871 | GTCTATGATTACAGACGAGAGGAATCAGGACAAACCAATCCCGTGGCTACCGAGCAGTAT | 3930 |
| QY | 441 | GlyThrValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHis | 460 |
| DB | 3931 | GGTTCTGTATCTCAACCTCCAGAGGCAACAGACAGCAGCTACCGCAGATGTCAAC | 3990 |
| QY | 461 | AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrrLeuGlnGlyPro | 480 |
| DB | 3991 | ACACAAGCGCTTCTTCAGGCGATGGTCTGGCAGCAGACAGATGTGTACTTTCAGGGGCC | 4050 |
| QY | 481 | IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly | 500 |
| DB | 4051 | ATCTGGGCAAGATTCCACACACGAGCGGACATTTTACCCCTCTCCCTCATGGGTGGA | 4110 |
| QY | 501 | PheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsn | 520 |
| DB | 4111 | TTGGAATTAACACCTCTCTCCAGANTTCTCATCAAGAACACCCCGTACCTGCGAAT | 4170 |
| QY | 521 | ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrrSerThrGly | 540 |
| DB | 4171 | CTTTCGACCACTTCAGTGGCGCAAGTTTCTCTTCTTCATCACACAGTACTTCCACGGGA | 4230 |
| QY | 541 | GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPro | 560 |
| DB | 4231 | CAGGTACGCGGGAGATCGAGTGGAGCTGAGAGGAAACAGCAACGCTGGTAATCCC | 4290 |
| QY | 561 | GluValGlnTyrrThrSerAsnTyrrAlaLysSerAlaAsnValAspPheThrValAspAsn | 580 |
| DB | 4291 | GAAATTCAGTACACTTCCAACTACAAAGCTGTGTTAATGTGGACTTTTACTGTGGACACT | 4350 |
| QY | 581 | AsnGlyLeuTyrrThrGluProArgProIleGlyThrArgTyrrLeuThrArgProLeu | 599 |
| DB | 4351 | AATGCGGTGATTACAGAGCTCGCCCATTTGGCCACAGATACCTGACTCGTAATCTG | 4407 |

XX AAT09008;
 AC
 XX
 DT
 XX
 XX
 DE
 XX
 KW Wild-type adeno-associated virus 2 genome.
 XX
 KW adeno-associated virus 2; AAV2; inverted terminal repeat; viral vector;
 KW DNA delivery; immunodeficiency virus protein; immunity; human; simian;
 KW neurological disorder; therapy; ss.
 XX
 XX Adeno-associated virus.
 OS
 XX
 XX W09534670-A2.
 FN
 XX
 XX 21-DEC-1995.
 PD
 XX
 XX 06-JUN-1995; 95WO-US071178.
 XX
 XX 06-JUN-1994; 94US-0254358.
 PR
 XX (CHIL-) CHILDRENS HOSPITAL INC.
 XX
 XX Johnson PR;
 XX
 XX WPI; 1996-049697/05.
 DR
 XX
 XX Recombinant adeno-associated virus genome contig. protein encoding
 PT DNA - flanked by inverted terminal repeats, for use in vaccines or
 PT for treatment of neuro-degenerative disease
 XX
 XX Example 1; Page 27-29; 44pp; English.
 PS
 XX The present sequence is that of the wild-type adeno-associated virus
 CC (AAV) 2, a replication-deficient parvovirus. cis-acting sequences
 CC directing viral DNA replication (ori), encapsidation/packaging (pkg) and
 CC host cell chromosome integration (int) are contained within the ITRs
 CC (inverted terminal repeats of 145 nucleotides). When AAV infects a
 CC human cell, the viral genome integrates into chromosome 19 resulting in
 CC latent infection of the cell. Prodn. of infectious virus does not occur
 CC unless the cell is infected with a helper virus (e.g., adenovirus or
 CC herpes-virus). AAV possesses unique features that make it attractive as
 CC a vector for delivering foreign DNA to cells. A vector including a
 CC recombinant AAV genome contig. a simian immunodeficiency virus (SIV) rev
 CC and envelop (gpi160) gene cassette was constructed from an existing plasmid
 CC designated psub201. The plasmid contains a modified wild-type AAV2
 CC genome. Specifically, two XbaI sites were added via linker addition at
 CC sequence positions 190 and 4484. These sites are internal to 191 bp ITRs
 CC which include the 145 bp ITRs of the AAV genome. The insertion of these
 CC sites allows the complete removal of the internal 4.3 kb fragment contig.
 CC the AAV rep-cap genes upon XbaI digestion of the plasmid. (See also
 CC AAT09009-10).
 XX
 SQ Sequence 4680 BP; 1198 A; 1265 C; 1254 G; 963 T; 0 other;

Alignment Scores:
 Pred. No.: 2,5e-215 Length: 4680
 Score: 2722.50 Matches: 492
 Percent Similarity: 89.00% Conservative: 42
 Best Local Similarity: 82.00% Mismatches: 64
 Query Match: 83.74% Indels: 2
 DB: 17 Gaps: 1

US-09-807-802A-15 (1-599) x AAT09008 (1-4680)

QY 1 ThrAlaProGlyLYeLysArgProValGluGlnSerProGlnGluProAspSerSerSer 20
 DB 2614 ACGGCTCCGGGAAAAAAGAGCCGGTAGACACTCTCTGTGGACCGCAGACTCTCTCTCG 2673

QY 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPhcGlyGlnThrGly 40
 DB 2674 GGAACCGGAAGCGCGGCAGCAGCTGCAAGAAAAAGATTGAATTTTGGTCAGACTGGA 2733

QY 41 AspSerGluSerValProAspProGlnProLeuGlyGluProAlaThrProAlaAla 60
 DB 2734 GACGCAGACTCAGTACTGACCCCGCTCTCGACAGCCAGCAGCCCGCTCTGGT 2793

QY 61 ValGlyProThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGlyGly 80
 DB 2794 CTGGGAACATAACGATGGCTACAGCAGTGGCGCACCATTGGCAGACATACGAGGGC 2853

QY 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
 DB 2854 GCCGACGGAGTGGGTAAATCTCCCGAAATGGCAATTGGCAATTCACATGGATGGCGAC 2913

QY 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTrpAsnAsnHisLeuTyr 120
 DB 2914 AGAGTCATACCCACAGCAGCCGAACTGGCGCCCTGCCACCTACAAACACCACTCTAC 2973

QY 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer 140
 DB 2974 AAACAAATTTCCAGCCCAATCA---GGAGCCTCGAAGCAGCAATCACTACTTTGGCTACAGC 3030

QY 141 ThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
 DB 3031 ACCCTTGGGGGTATTTTGAATTCACAGATTCCACTGCCACTTTTCCACACGCTGACTGG 3090

QY 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
 DB 3091 CAAAGACTCATCAACAACTCGGGATTCGAGCCCAAGAGACTCAACTTCAAGCTCTTT 3150

QY 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeu 200
 DB 3151 AACATTCAGTCAAGAGGTCACGAGATGACGGTACGAGCAGCATTTGCCAATAACCTT 3210

QY 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
 DB 3211 ACCAGACGGTTTCAAGTGTATTACTGACTCGAGTACCGACTCCGCTCGGCTCG 3270

QY 221 AlaHisGlnGlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGly 240
 DB 3271 GCGCATCAAGAGTGCCTCCCGCGTTCACAGCAGACGTCTTCATGGTGCCACATATGGA 3330

QY 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
 DB 3331 TACTTCACTTCAACAAACGAGTACGAGCAGTAGGACGCTCTTCAATTTTACTGCTGGAG 3390

QY 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
 DB 3391 TACTTCTCTTCTCAGATGCTGCGTACCGGAAACAACATTTTACCTTCAGCTACACTTTTGAG 3450

QY 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
 DB 3451 GAGCTTCTTTCACAGCAGCTACGCTCACAGCCAGAGTCTGGACCGCTCTCATGAATCCT 3510

QY 301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
 DB 3511 CTCAATCGACCACTACCTGTATTACTTGGCAGAGCAACAACACTCCCAAGTGGAAACCAAC 3570

QY 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
 DB 3571 CAGTCAAGGGCTTCAAGTTTCTCAGCGCGGAGCGAGTACCATTCGGGACCACTCTAGGAAC 3630

QY 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
 DB 3631 TGGCTTCTGGACCTCTGTATCCCGCCAGCGAGTATCAAGACATCTCGCGGATACACAC 3690

QY 361 AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 380
 DB 3691 AACAGTGAATACTCGTGGACTGGAGCTACCAAGTACCACTCAATGGCAGAGACTCTCTG 3750

QY 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet 400
 DB 3751 GTGAATCCGGGGCCCGCCATGGCAAGCCACAAAGACGATGAAGAAAAAGTTTTTCTCTCAG 3810

QY 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: February 14, 2004, 16:11:58 ; Search time 88.4559 Seconds
(without alignments)
2988.931 Million cell updates/sec

Title: US-09-807-802A-15
Perfect score: 3251
Sequence: 1 TAPGKRPVQSPQDPSSS.....NNGLYTEPRICTRYLTRPL 599

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=xlh
-Q=/cgn2_1/usfto_spool/US09807802/runat_11022004_175609_15955/app_query.fasta.1.2389
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09807802 @CNC 1.1.131 @runat_11022004_175609_15955 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------------|
| 1 | 2801 | 86.2 | 8179 | 4 | US-09-438-268-5 |
| 2 | 2764.5 | 85.0 | 4072 | 4 | US-09-770-315-4 |
| 3 | 2764.5 | 85.0 | 7557 | 4 | US-09-770-315-3 |
| 4 | 2764.5 | 85.0 | 8658 | 4 | US-09-770-315-2 |
| 5 | 2722.5 | 83.7 | 4680 | 1 | US-08-254-358-1 |
| 6 | 2722.5 | 83.7 | 4680 | 1 | US-08-475-391-1 |
| 7 | 2722.5 | 83.7 | 4680 | 2 | US-08-709-609-1 |
| 8 | 2722.5 | 83.7 | 4680 | 5 | PCT-US95-07178-1 |
| 9 | 2398.5 | 73.8 | 8151 | 4 | US-09-438-268-2 |
| 10 | 1830.5 | 56.3 | 1800 | 4 | US-09-532-594B-17 |
| 11 | 1830.5 | 56.3 | 2208 | 4 | US-09-532-594B-5 |
| 12 | 1830.5 | 56.3 | 4767 | 4 | US-09-532-594B-1 |

| | | | | | | | |
|---|----|--------|------|---------|---|----------------------|--------------------|
| c | 13 | 1830.5 | 56.3 | 7214 | 4 | US-09-438-268-1 | Sequence 1, Appli |
| | 14 | 1690.5 | 52.0 | 1617 | 4 | US-09-532-594B-19 | Sequence 19, Appl |
| | 15 | 772 | 23.7 | 2271 | 4 | US-09-438-268-3 | Sequence 3, Appli |
| | 16 | 247 | 7.6 | 1740 | 1 | US-07-969-213-1 | Sequence 1, Appli |
| | 17 | 231.5 | 7.1 | 2062 | 6 | 5223424-3 | Patent No. 5223424 |
| | 18 | 220 | 6.8 | 1820 | 6 | 5223424-12 | Patent No. 5223424 |
| | 19 | 209 | 6.4 | 5049 | 1 | US-08-336-345-1 | Sequence 1, Appli |
| | 20 | 209 | 6.4 | 5049 | 1 | US-08-336-345-2 | Sequence 2, Appli |
| | 21 | 209 | 6.4 | 5049 | 2 | US-08-647-655-1 | Sequence 1, Appli |
| | 22 | 209 | 6.4 | 5049 | 2 | US-08-647-655-2 | Sequence 2, Appli |
| | 23 | 203 | 6.2 | 2254 | 3 | US-08-552-369-1 | Sequence 1, Appli |
| | 24 | 192 | 5.9 | 1752 | 3 | US-09-022-949-1 | Sequence 1, Appli |
| | 25 | 184 | 5.7 | 1755 | 2 | US-08-317-785-1 | Sequence 1, Appli |
| c | 26 | 131 | 4.0 | 4188 | 4 | US-09-252-991A-13774 | Sequence 13774, A |
| | 27 | 131 | 4.0 | 8211 | 4 | US-09-252-991A-13656 | Sequence 13656, A |
| c | 28 | 123.5 | 3.8 | 1896 | 2 | US-08-198-446B-8 | Sequence 8, Appli |
| c | 29 | 123.5 | 3.8 | 1896 | 2 | US-08-870-693-8 | Sequence 8, Appli |
| c | 30 | 120 | 3.7 | 4403765 | 3 | US-09-103-840A-2 | Sequence 2, Appli |
| c | 31 | 119.5 | 3.7 | 4411529 | 3 | US-09-103-840A-1 | Sequence 1, Appli |
| c | 32 | 119 | 3.7 | 3288 | 4 | US-09-107-532A-201 | Sequence 201, App |
| c | 33 | 119 | 3.7 | 4214 | 4 | US-09-122-135-1 | Sequence 1, Appli |
| c | 34 | 117.5 | 3.6 | 3906 | 4 | US-09-996-243-83 | Sequence 83, Appl |
| | 35 | 114 | 3.5 | 3155 | 2 | US-08-231-193A-13 | Sequence 13, Appl |
| | 36 | 114 | 3.5 | 3155 | 2 | US-08-486-273A-13 | Sequence 13, Appl |
| | 37 | 114 | 3.5 | 3155 | 3 | US-08-480-474-13 | Sequence 13, Appl |
| | 38 | 114 | 3.5 | 3155 | 3 | US-08-940-086A-13 | Sequence 13, Appl |
| | 39 | 114 | 3.5 | 3155 | 4 | US-08-940-035A-13 | Sequence 13, Appl |
| | 40 | 114 | 3.5 | 3155 | 4 | US-08-935-105A-13 | Sequence 13, Appl |
| | 41 | 114 | 3.5 | 3155 | 4 | US-09-648-797-13 | Sequence 13, Appl |
| | 42 | 114 | 3.5 | 3155 | 4 | US-09-386-123-13 | Sequence 13, Appl |
| | 43 | 114 | 3.5 | 3211 | 2 | US-08-231-193A-21 | Sequence 21, Appl |
| | 44 | 114 | 3.5 | 3211 | 2 | US-08-486-273A-21 | Sequence 21, Appl |
| | 45 | 114 | 3.5 | 3211 | 3 | US-08-480-474-21 | Sequence 21, Appl |

ALIGNMENTS

RESULT 1
US-09-438-268-5
; Sequence 5, Application US/09438268
; Patent No. 6491907

; GENERAL INFORMATION:
; APPLICANT: Rabinowitz, Joseph E.
; APPLICANT: Samulski, Richard J.
; APPLICANT: Xiao, Weidong
; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 5470-186
; CURRENT APPLICATION NUMBER: US/09/438,268
; EARLIER FILING DATE: 1999-11-10
; EARLIER APPLICATION NUMBER: 60/107,840
; EARLIER FILING DATE: 1998-11-10
; EARLIER APPLICATION NUMBER: 60/123,651
; EARLIER FILING DATE: 1999-03-10
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 8179
; TYPE: DNA
; ORGANISM: Virus
US-09-438-268-5

Alignment Scores:
Pred. No.: 8,598-261 Length: 8179
Score: 2801.00 Matches: 509
Percent Similarity: 90.33% Conservative: 33
Best Local Similarity: 84.83% Mismatches: 56
Query Match: 86.16% Indels: 2
DB: 4 Gaps: 2

US-09-807-802A-15 (1-599) x US-09-438-268-5 (1-8179)

QY 1 ThrAlaProGlyLysArgProValGluGlnSerProGlnGluProAspSerSer 20

2544 ACAGCTCCGGGAAAAAGAGCGCGTAGAGCACTCTCTGTGGACCGAGACTCTCTCTCG 2603
21 GlyIleGlyLeuThrGlyGlnGlnProAlaLysArgLeuAsnPheGlyGlnThrGly 40
2604 GGAACCGGAAAGCGCGCGCGAGCGCTGCAAGAAAGAGATTGAAATTTGGTCAGACTGGA 2663
41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaala 60
2664 GAGCAGACTCAGTACCTGACCCCGAGCGCTCTCGACAGCCACCGAGCGCCCTCTCGT 2723
61 ValGlyProThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAspAsnGluGly 80
2724 CTGGGNACTAATACGATGCTACAGCGAGTGGCGGACCAATGCGCAGACATACAGCGGGC 2783
81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
2784 GCGCAGGAGTGGTAAATTCCTCCGGAATTCGCAATTCGCAATTCACATGATGATGGCGGAC 2843
101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrThrAsnAsnHisLeuTyr 120
2844 AGAGTCATCACCACGACGACCCGAACTTGGGCGCTGCCACCTCAACAAACACCTCTAC 2903
121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer 140
2904 AAACAAATTTCCAGCCCAATCA--GGAGCTCGACGACCAATCACTACTTTGGCTACAGC 2960
141 ThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
2961 ACCCTTGGGGGTATTTTACATTTCAACAGATTCACAGTTCACCTTTTCCACAGTCACTGG 3020
161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
3021 CAAAGACTCATCAACAAACAACTGGGATTCGAGCCCAAGAGACTCAACTTCAAGTCTTT 3080
181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeu 200
3081 AACATTCAGTCAACAGAGTCAAGCGATGAGCGGTACGACGAGATTCGCAATCAACTTT 3140
201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
3141 ACCAGACGGTTCAGGTGTTTACTGACTCGAGTACAGCTCCCGTACGTCTCGGGTGC 3200
221 AlaHisGlnGlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGly 240
3201 GCGCAACAGGCTGTCTCCCGCGTTCAGCGGAGCTTCTCATGTCTCTCATGTATGA 3260
241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
3261 TACCTCACCCTGAACACGAGTCAAGCGGTGGGACGCTCATCTTTTACTGCGTGGAG 3320
261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
3321 TACTTCCCTTCGAGATGCTAAGGACTGGAAATTAACCTTCAATTCAGTATACCTTCGAG 3380
281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
3381 GATGTACTTTTACAGCAGCTACGCTCAGCGCAGGTTGGATCGCTTGATGATCTCT 3440
301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGln--AsnGlnSerGlySerAla 319
3441 CTTATTGATCAGTATCTGTACTACCTGAAACAGAACGCAAGGAAACCTCTGGAACAAAC 3500
320 GlnAsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLys 339
3501 AACCAATCAGCGTGTCTTTTACGCGAGCGTGGCGCTCAGTCTATGCTTTTGGAGCGCAGA 3560
340 AsnTrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsn 359
3561 AATTTGGTACTTGGGCTCTGCTACCGGCAACAGAGACTTTCAGAGACTGCTACAGACAC 3620
360 AsnAsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSer 379

3621 AACCAAGTAACTTTCTTGGACAGCGCGCAGCAAAATATCATCTCAATGGCGCGACTCG 3680
380 IleIleLeuProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePhePro 399
3681 CTGTGTAATCCAGGACCGAGCTATGCCAGTACCAAGGACGATGAAGAAAAATTTTCCCT 3740
400 MetSerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAsp 419
3741 ATGCACGGCAATCTAATATTTGGCAAGAGGCAACGGCAAGTACGAGCAATTAGAT 3800
420 AsnValMetIleThrAspGluGluLeuLysAlaThrAsnProValAlaThrGluArg 439
3801 AATGTAATGATTAACGATGAAGAGATTCGTACCAACCAATCTCTGTGGCAACAGAGAG 3860
440 PheGlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspVal 459
3861 TATGGAATCTGGCAAAATAACTTCAGAGCTCAATACAGCTCCACAGCTGGAACTGTC 3920
460 HisAlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGly 479
3921 AATCATCAGGGGCTTTACCTGGCATGGTGTGGCAAGATCGTGACGTGCTACTTCAAGGA 3980
480 ProIleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGly 499
3981 CCTATCTGGCAAAAGATTCCTCACAAGGATGGACACTTTTCATCTCTCTCTCTGATGGGA 4040
500 GlyPheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAla 519
4041 GGTCTTGGAGTGAACATCCGCTCCTCAATCATGATCAAAAATATCTCCGGTACTCTGG 4100
520 AsnProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThr 539
4101 AATCTCTCGACCACTTCAGTGGCGCAAGTTTGTCTCTTCATCACACAGTACTCCACG 4160
540 GlyGlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsn 559
4161 GCACAGGTCAGCGTGGAGATCGAGTGGAGCTGCAGAGAGGAAACAGCAACGCTGGAAAT 4220
560 ProGluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAsp 579
4221 CCGAAATTCAGTACACTTCCAACTACCAAGTCTGTATTCGTGGAGTACTCCGTGGAT 4280
580 AsnAsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
4281 ACTAATGGGTGTTTTCAGAGCTCGCCCATTCGACAGATACCTGACTCGTAATCTG 4340

RESULT 2

US-09-770-315-4
; Sequence 4, Application US/09770315
; Patent No. 6429001
; GENERAL INFORMATION:
; APPLICANT: Chiron Corporation
; TITLE OF INVENTION: Recombinant AAV Packaging Systems
; FILE REFERENCE: 20263-501
; CURRENT APPLICATION NUMBER: US/09/770,315
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,536
; PRIOR FILING DATE: 2000-01-26
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 4072
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: recombinant DNA
US-09-770-315-4

Alignment Scores:
Pred. No.: 9,78e-258 Length: 4072
Score: 2764.50 Matches: 496
Percent Similarity: 89.98% Conservative: 43
Best Local Similarity: 82.80% Mismatches: 59

Query Match: 85.04% Indels: 1
DB: 4 Gaps: 1
US-09-807-802a-15 (1-599) x US-09-770-315-4 (1-4072)

QY 1 ThrAlaProGlyValylsArgProValGluGlnSerProGlnGluProAspSerSerSer 20
DB 1895 ACGGCTCCGGGAAAAGAGGCGGTAGACACTCTCTGTGTGGAGCCAGACTCCCTCTCG 1954

QY 21 GlyIleGlyThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGly 40
DB 1955 GGAACCGGNAAGCGGCGCAGCAGCTGCGAAGAAAGATTGAAATTTGGTCAGACTGGA 2014

QY 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
DB 2015 GACGACAGACTAGTACTGACCCCGACGCTCTCGGACAGCCACAGCCGCCCCCTCTGT 2074

QY 61 ValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGly 80
DB 2075 CTGGGAACCTAATACGATGGCTACAGGCAGTGGCGCACCACCAATGGCAGACAATAACGAGGCG 2134

QY 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
DB 2135 GCGGACGGAGTGGTAATCTCTCGGAAATTTGGCAATTCGATTCACATGGATGGCGCAC 2194

QY 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrThrAsnAsnHisLeuTyr 120
DB 2195 AGAGTCATCACCCAGCAGCCGGAACCTGGGCGCTGCCACCTCAACAACACCCCTCTAC 2254

QY 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer 140
DB 2255 AAACAAATTTCCAGGCAATCA---GGAGCTCGAAGCAACATCACTACTCTTTGGCTACAGC 2311

QY 141 ThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
DB 2312 ACCCTTGGGGTATTTTGACTTCAACAGATTCCACTGCCACTTTTCCACGAGTACTGG 2371

QY 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
DB 2372 CAAGACTCATCAACAACACTGGGGATTCCGACCCCAAGAGACTCAACTTCAAGCTCTTT 2431

QY 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrIleAsnAsnLeu 200
DB 2432 AACATTCAAGTCAAGAGGTCAAGCAGATGACGGTACGACGAGATTGCCAATTAACCTT 2491

QY 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
DB 2492 ACCAGCAGGTTTCAGGTCTTTACTGACTCGGAGTACCAAGCTCCCGTACGTCCTCGGCTCG 2551

QY 221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGly 240
DB 2552 GCGCATCAAGGATGCTCCCGCGTTCCCGCAGCAGACGCTTTCATGGTGCCACAGTATGGA 2611

QY 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
DB 2612 TACTTCAACCTCGAACACGCGGAGTCAGGAGTAGGACGCTCTTCAATTTACTGCTGGAG 2671

QY 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
DB 2672 TACTTCTCTCAGATGCTGGTACCGGAAACACTTTTACTTTCAGTACACTTTTGAG 2731

QY 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
DB 2732 GACGTTCTTCCACAGCAGCTAGCTCAGCAGCAGAGTCTGGACCGTCTCATGAATCCT 2791

QY 301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
DB 2792 CTATCGACCACTAGTACTGTATTACTTTGACGAGAAACAAACACTCCAAAGTGAACCAACG 2851

QY 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
DB 2852 CAGTCAAGGCTTCAGTTTTTCTCAGCGCGGAGGTGACATTCGGGACCACTCTAGGNAC 2911

QY 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
DB 2912 TGGCTTCTGGACCTGTTACCGCCAGCAGGAGTATCAAGACATCTCGCGATACCAAC 2971

QY 361 AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 380
DB 2972 AACAGTGAATACTCGTGACTGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTG 3031

QY 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet 400
DB 3032 GTGAATCCGGGCCCGCCCATGGCAAGCCACAGGACGATGAAGAAAAGTTTTTCTCTCAG 3091

QY 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
DB 3092 AGCGGGTTCTCATCTTTGGGAACAGGCTCAGAGAAAACAAATGTGGACATTTGAAAAG 3151

QY 421 ValMetIleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
DB 3152 GTCATGATTACAGACGAGGAAATCAGGACCAACCAATCCCGTGGCTACGAGCAGTAT 3211

QY 441 GlyThrValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHis 460
DB 3212 GGTTCGTGATCTCAACACTCCAGAGAGGCAACAGACGAGCTACCGCAGATGTCAAC 3271

QY 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro 480
DB 3272 ACACAAGCGTTCTTCAGGCGATGTTCTGGCAGGACAGAGATGTGTACCTTCAGGGGCC 3331

QY 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
DB 3332 ATCTGGGCAAGATTCCACACAGCAGCGACATTTTCCACCCCTCTCCCTCATGGTGA 3391

QY 501 PheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsn 520
DB 3392 TTCGACTTAAACACCTCTCCACAGATTCTCATCAAGAAACACCCCGTACCTGCGAAT 3451

QY 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
DB 3452 CCTTCGACCACTTCAGTGGCGCAAGTTTGTCTTCTTCATCACACAGTACTCCACGGA 3511

QY 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPro 560
DB 3512 CAGTTCAGCTGGAGATCGAGTGGAGCTCAGAAAGGAAACAGCAACCGTGGAAATCCC 3571

QY 561 GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsn 580
DB 3572 GAAATTCAGTACACTTCCAACTACACAGTCTGTTAATGTGGACTTTTACTGTGGACACT 3631

QY 581 AsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
DB 3632 AATGGCGTGATTTCAGAGCCTCGCCCATTTGGCACCAGATACCTGACTCGTAATCTG 3688

RESULT 3
US-09-770-315-3
; Sequence 3, Application US/09770315
; Patent No. 6429001
; GENERAL INFORMATION:
; APPLICANT: Chiron Corporation
; TITLE OF INVENTION: Recombinant AAV Packaging Systems
; FILE REFERENCE: 20263-501
; CURRENT APPLICATION NUMBER: US/09/770,315
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,536
; PRIOR FILING DATE: 2000-01-26
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 7557
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: recombinant DNA
; US-09-770-315-3

Alignment Scores:

Pred. No.: 2,62e-257 Length: 7557
 Score: 2764.50 Matches: 496
 Percent Similarity: 89.98% Conservative: 43
 Best Local Similarity: 82.80% Mismatches: 59
 Query Match: 85.04% Indels: 1
 DB: 4 Gaps: 1

US-09-807-802A-15 (1-599) x US-09-770-315-3 (1-7557)

QY 1 ThrAlaProGlyValSerArgProValGluGlnSerProGlnGluProAspSerSer 20
 DB 2642 ACCGCTCCGGGAAAAGAGCGCGTAGACACTCTCTGTGGAGCGAGACTCTCTCTG 2701
 QY 21 GlyIleGlyThrGlyGlnGlnProAlaLysArgLeuAenPheGlyGlnThrGly 40
 DB 2702 GGAACCGGAAGCGGCGCAGCAGCTGCAAGAAAAGATTGAATTTGGTCAGACTGGA 2761
 QY 41 AspSerGluSerValProAspProGlnProLeuGluProProAlaThrProAlaAla 60
 DB 2762 GACGCGAGACTCAGTACTGACCCCGCAGCTCTCGGACAGCCACCGAGCGCCCTCTGGT 2821
 QY 61 ValGlyProThrThrMetAlaSerGlyGlyGlyValAlaProMetAlaAspAenGluGly 80
 DB 2822 CTGGGAACCTAATACGATGGCTACAGGAGTGGCGCACCAATGSCAGACATACAGAGGC 2881
 QY 81 AlaAspGlyValGlyAenAlaSerGlyValSerHisCysAspSerThrTrpLeuGlyAsp 100
 DB 2882 GCCGACGGAGTGGTAAATCTCCGGGAAATGGCAATGGCATTCACATGGATGGCGGAC 2941
 QY 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTrpAenAenHisLeuTyr 120
 DB 2942 AGAGTCATACACCAACGACCCGAACTGGGCGCTGCGCCACCTACAAACACCACTCTAC 3001
 QY 121 LysGlnIleSerSerAlaSerThrGlyValSerAenAspAenHisTyrPheGlyTyrSer 140
 DB 3002 AACAAATTTCCAGCCATCA--GGAGCCTCGAACGACATCACTACTTGGCTACAGC 3058
 QY 141 ThrProTrpGlyTyrPheAspPheAenArgPheHisCysHisPheSerProArgAspTrp 160
 DB 3059 ACCCTTGGGGTATTTTGACTTCAACAGATTCCACTGCCACTTTTCCACCACGTGACTG 3118
 QY 161 GluArgLeuIleAenAenAenTrpGlyPheArgProLysArgLeuAenPheLysLeuPhe 180
 DB 3119 CAAGACTCATCAACAACACTGGGATTCGACCCCAAGAGACTCACTTCAAGCTCTTT 3178
 QY 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAenAenLeu 200
 DB 3179 AACATTCAAGTCAAGAGGTCACGCAGATGACGGTACGAGCAGATTGCCAATACCTT 3238
 QY 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
 DB 3239 ACCAGCAGGTTTACGGTGTCTTACTGACTCGGAGTACCAGCTCCCGTACGTCTCGGCTG 3298
 QY 221 AlaHisGlnGlyCysValLeuProPheProAlaAspValPheMetIleProGlnTyrGly 240
 DB 3299 GCGCATCAGATGCTCTCCCGCTTCCAGCAGAGCTCTTCATGTGGCCACAGTATGGA 3358
 QY 241 TyrLeuThrLeuAenAenGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
 DB 3359 TACTCTACCTGACAAACGGGAGTCAGGAGTAGGAGCTCTTCAATTTACTGCCTGGAG 3418
 QY 261 TyrPheProSerGlnMetLeuArgThrGlyAenAenPheThrPheSerTyrThrPheGlu 280
 DB 3419 TACTTTCTCTCAGATGCTCGGTACCGGAAACAACTTTTACCTTCAGCTACACTTTTGG 3478
 QY 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAenPro 300
 DB 3479 GAGCTTCTTCCACAGCAGCTACGCTCAGCCAGAGTCTGGACCGTCTCATGAATCT 3538
 QY 301 LeuIleAspGlnTyrLeuTyrTyrLeuAenArgThrGlnAenGlnSerGlySerAlaGln 320
 DB 3201 LeuIleAspGlnTyrLeuTyrTyrLeuAenArgThrGlnAenGlnSerGlySerAlaGln 320

DB 3539 CTCATCGACCACTCTGTTACTTGTAGCAGAACAAACACTCCCAAGTGGACCAACACG 3598
 QY 321 AenLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAen 340
 DB 3599 CAGTCAAGGCTTCACTTTCTCAGCGCGGAGCGAGTGCATTCGGGACCAAGTCTAGAAC 3658
 QY 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAenAen 360
 DB 3659 TGGCTTCTCGGACCTGTTTACCGCAGCAGCAGATCAAGACATCTCGGATCAACAC 3718
 QY 361 AenSerAenPheThrTrpThrGlyAlaSerLysTyrAenLeuAenGlnGluSerIle 380
 DB 3719 AACAGTGAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3778
 QY 381 IleAenProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet 400
 DB 3779 GTGAATCCGGGCGCGCCATGCGCAAGCCACAGGACGATGAAGAAAAGTTTTTCTCTCAG 3838
 QY 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAenThrAlaLeuAenAen 420
 DB 3839 ACCGGGTTCTCTCTTTGGGAGCAGGCTCAGAGAAAACAATGTGGACATTTGAAAG 3898
 QY 421 ValMetIleThrAspGluGluGluLysAlaThrAenProValAlaThrGluArgPhe 440
 DB 3899 GTCATGATTACAGACGAAAGAGAAATCAGACCAACCAATCCCGTGGCTACGGAGCAGTAT 3958
 QY 441 GlyThrValAlaValAenPheGlnSerSerSerThrAspProAlaThrGlyAspValHis 460
 DB 3959 GGTCTGTATCTACCAACCTCCAGAGAGGCAACAGACAGCTACCGCAGATGTCAAC 4018
 QY 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro 480
 DB 4019 ACACAGGGCTTCTCCAGGCATGCTTGGCAGGACAGAGATGTGTACCTTCAGGGGCC 4078
 QY 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
 DB 4079 ATCTGGCAAGATTTCCACACCGCAGCAGATTTTCACTCCCTCTCCCTCATGGGTGA 4138
 QY 501 PheGlyLeuLysAenProProGlnIleLeuIleLysAenThrProValProAlaAen 520
 DB 4139 TTCCGACTTAAACACCTCTCCAGATTTCTCATCAAGAACACCCCGGTACTCTCGGAAT 4198
 QY 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
 DB 4199 CTTTCGACCACTTCAGTGGGCAAGTTTGTCTCTTCATCACACAGTACTCCACGGA 4258
 QY 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAenSerLysArgTrpAenPro 560
 DB 4259 CAGGTCAGCGTGGAGATCGAGTGGGAGCTGCAGAGGAAAGAAACAGCAACCTGGAATCCC 4318
 QY 561 GluValGlnTyrThrSerAenTyrAlaLysSerAlaAenValAenPheThrValAenAen 580
 DB 4319 GAAATTCAGTACACTTCCAACTACAAAGTCTGTAAATGTGGACTTTACTGTGGACACT 4378
 QY 581 AsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
 DB 4379 AATGGCTGTATTCAGAGCTCGCCCATTTGGCAGCAGATACCTGACTCGTAATCTG 4435

RESULT 4

US-09-770-315-2
 ; Sequence 2, Application US/09770315
 ; Patent No. 6429001
 ; GENERAL INFORMATION:
 ; APPLICANT: Chiron Corporation
 ; TITLE OF INVENTION: Recombinant AAV Packaging Systems
 ; FILE REFERENCE: 20263-501
 ; CURRENT APPLICATION NUMBER: US/09/770,315
 ; CURRENT FILING DATE: 2001-01-26
 ; PRIOR APPLICATION NUMBER: US 60/178,536
 ; PRIOR FILING DATE: 2000-01-26
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 2

; LENGTH: 8698

; TYPE: DNA

; ORGANISM: Unknown

; FEATURE:

; OTHER INFORMATION: recombinant DNA

; US-09-770-315-2

Alignment Scores:

Pred. No.: 3 286-257 Length: 8698
 Score: 2764.50 Matches: 496
 Percent Similarity: 89.98% Conservative: 43
 Best Local Similarity: 82.80% Mismatches: 59
 Query Match: 85.04% Indels: 1
 DB: 4 Gaps: 1

US-09-807-802a-15 (1-599) x US-09-770-315-2 (1-8698)

QY 1 ThrAlaProGlyLysLysArgProValGluInSerProGlnProAspSerSerSer 20
 DB 2614 ACGGCTCCGGGAAAAAGAGCGGTAGAGCACTCTCTGTGGAGCCAGACTCTCTCTCG 2673
 QY 21 GlyLeGlyThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGly 40
 DB 2674 GGAACCGGAAGCGCGGCAGAGCCTGCAAGAAAAAGATTGAATTTTGGTCAGACTGGA 2733
 QY 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
 DB 2734 GAGCGAGACTAGTACTGACCCCGAGCCTCTCGAGCCAGCCAGCAGCCCTCTGCT 2793
 QY 61 ValGlyProThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAspAsnGluGly 80
 DB 2794 CTGGGAACCTAATACGATGGCTACAGGCAGTGGCGCCACCAATGGCAGACAATAACAGGGC 2853
 QY 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
 DB 2854 GCCGACGAGTGGGTAAATCTCCGGGAAATGGCATTCGGATTCCAGTGGATGGCGGAC 2913
 QY 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTrpAsnAsnHisLeuTyr 120
 DB 2914 AGAGTCATCACCAACAGCAGCCGAGCTGGGCTGCCCCACTTACCAACACCACTCTAC 2973
 QY 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisThrPheGlyTyrSer 140
 DB 2974 AAACAAATTTCCAGCCCAATCA--GGAGCCTCGAAGCAGCAATCACTACTTTGGCTACAGC 3030
 QY 141 ThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
 DB 3031 ACCCTTTGGGGGTATTTTGACTTCAACAGATTCCACTGCCCACTTTTCCACCACTGACTGG 3090
 QY 161 GlnArgLeuIleAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
 DB 3091 CNAAGACTCATCAACACACTGGGATTCGGACCCCAAGAGACTCAACTTCAAGCTCTTT 3150
 QY 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeu 200
 DB 3151 AACATTCAAGTCAAGAGGTCAAGCATGACGGTACGAGCAGCATTTGCCAATAACCTT 3210
 QY 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
 DB 3211 ACCAGCAGGTTCAGGTGTTTACTTGACTTCGAGTACCAGCTCCCGTACGTCCTCGGCTCG 3270
 QY 221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGly 240
 DB 3271 GCGCATCAAGATGCTCCCGCTCCAGCAGAGCTCTTCATGGTGGCCACAGTATGGA 3330
 QY 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
 DB 3331 TACCTCACCTGTAAACAACGGAGTCAGCAGTAGGACGCTCTTCACTTTTACTGCCTGGAG 3390
 QY 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
 DB 3391 TACTTTCTCTCAGATGCTGCTGCTACCGGAAACAACCTTTACCTTCAGCTACACTTTTGAG 3450

QY 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
 DB 3451 GACGTTCTTTCCACAGCAGCTACGCTCACAGCCAGAGTCTGGACCGTCTCATGAATCCT 3510
 QY 301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
 DB 3511 CTCATCGACCACTGATCTGATTACTTGAGCAGAAACAACACTCCAAGTGGAAACCAACACG 3570
 QY 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
 DB 3571 CAGTCAAGGCTTCAGTTTCTCAGCCCGGAGCTAGATGACATTCGGGACCACTTAGGAAC 3630
 QY 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
 DB 3631 TGGCTTCTCGACCTGTTTACCGCAGCAGCAGATATCAAGAGACTCTGCGGATAACAAC 3690
 QY 361 AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 380
 DB 3691 AACAGTGAATATCTCGTGACTGGAGCTACCAAGTACCACTCAATGGCAGAGACTCTCTG 3750
 QY 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet 400
 DB 3751 GTGATCTCCGGCCCGCCATGCGCAAGCCACCAAGGACGATGAAGAAAGTCTTCTCTCAG 3810
 QY 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
 DB 3811 AGCGGGTCTCATCTTTGGGAAGCAAGGCTCAGAGAAACAATAATGTGGACATTTGAAAG 3870
 QY 421 ValMetIleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
 DB 3871 GTCATGATTACAGCAAGAGGAATCAGGCAACCAATCCGCTGGCTACGAGCGATAT 3930
 QY 441 GlyThrValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHis 460
 DB 3931 GGTCTGTATCTTACCAACTCCAGAGAGGCAACAGACAGCTACCGCAGATGTCAAC 3990
 QY 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro 480
 DB 3991 ACACAAGCGGTCTTCCAGGCATGCTCTGGCAGGACAGAGATGTGTACTTCCAGGGGCC 4050
 QY 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
 DB 4051 ATCTGGGCAAGATTCCACACGACGAGCAATTTTCACTCCCTCTCCCTCTCATGGTGA 4110
 QY 501 PheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsn 520
 DB 4111 TTCGNCITTAACACCTCTCCACAGATTCTCATCAAGAACACCCCGTACTTCCGAAT 4170
 QY 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
 DB 4171 CCTTCGACCCCTTCAGTCCGCAAGTTTGTCTTCTCATCACACAGTACTTCCAGGGA 4230
 QY 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPro 560
 DB 4231 CAGGTGAGGTGGAGTCCAGTGGGAGCTGCAGAGAGAAACAGCAACCGCTGGATCCC 4290
 QY 561 GluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsn 580
 DB 4291 GAAATTCAGTACACTTCCCACTACCAAGTCTGTTAATGTGGACTTTTACTTGGACACT 4350
 QY 581 AsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
 DB 4351 AATGGCGTGTATTACAGCGCTCGCCCCATTGGCAGCATACCTGACTGCTAATCTG 4407

RESULT 5

US-08-254-358-1

; Sequence 1, Application US/08254358

; Patent No. 5658785

; GENERAL INFORMATION:

; APPLICANT: Johnson, Philip R.

; TITLE OF INVENTION: Adeno-Associated Virus Materials and

; TITLE OF INVENTION: Methods

; NUMBER OF SEQUENCES: 3

;; CORRESPONDENCE ADDRESS:
 ;; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ;; STREET: 6300 Sears Tower, 233 S. Wacker Drive
 ;; CITY: Chicago
 ;; STATE: Illinois
 ;; COUNTRY: USA
 ;; ZIP: 60606

;; COMPUTER READABLE FORM:
 ;; MEDIUM TYPE: Floppy disk
 ;; COMPUTER: IBM PC compatible
 ;; OPERATING SYSTEM: PC-DOS/MS-DOS
 ;; SOFTWARE: Patent in Release #1.0, Version #1.25
 ;; CURRENT APPLICATION DATA:
 ;; APPLICATION NUMBER: US/08/254,358

;; FILING DATE:

;; CLASSIFICATION: 435

;; ATTORNEY/AGENT INFORMATION:

;; NAME: No. 5658785and, Greta E.

;; REGISTRATION NUMBER: 35,302

;; REFERENCE/DOCKET NUMBER: 31975

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (312) 474-6300

;; TELEFAX: (312) 474-0448

;; TELEX: 25-3856

;; INFORMATION FOR SEQ ID NO: 1:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 4680 base pairs

;; TYPE: nucleic acid

;; STRANDEDNESS: single

;; TOPOLOGY: linear

;; MOLECULE TYPE: DNA (genomic)

;; US-08-254-358-1

Alignment Scores:

Pred. No.: 1.44e-253 Length: 4680
 Score: 2722.50 Matches: 492
 Percent Similarity: 89.00% Conservative: 42
 Best Local Similarity: 82.00% Mismatches: 64
 Query Match: 83.74% Indels: 2
 DB: 1 Gaps: 1

US-09-807-802A-15 (1-599) x US-08-254-358-1 (1-4680)

QY 1 ThrAlaProGlyLysArgProValGluGlnSerProGlnGluProAspSerSerSer 20
 DB 2614 ACGGCTCCGGGAAAAAGAGCCGGTAGACACTCTCTCTGGAGCCAGACTCTCTCTCG 2673
 QY 21 GlyTleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGly 40
 DB 2674 GGAACCGGAAGCGCGGCAGACGCTGCAAGAAAGATTGAATTTGGTCAGACTGGA 2733
 QY 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
 DB 2734 GACGAGACTCAGTACCTGACCCCGCCAGCCTCTCGGACAGCCAGCCAGCCCTCTGGT 2793
 QY 61 ValGlyProThrThrMetAlaSerClyGlyAlaProMetAlaAspAsnAsnGlyGly 80
 DB 2794 CTGGGAACCTAATACGATGGCTACAGCGAGTGCGGCCCAATGGCAGACAAATACGAGGC 2853
 QY 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
 DB 2854 GCCGACGAGTGGGTAAATCTCCGGAATTTGGCATTTGGCATTTCCACATGGATGGCGAC 2913
 QY 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrThrAsnAsnHisLeuTyr 120
 DB 2914 AGAGTCATCACCCAGCAGCCGACCTGGGCGCTTGGCCACCTGACCAACACCACTCTAC 2973
 QY 121 LysGlnIleSerSerAlaSerThrClyValaSerAsnAspAsnHisTyrPheGlyTyrSer 140
 DB 2974 AAACAAATTTCCAGCCCAATCA--GGAGCCTCGAACGACAAATCACTACTTTGGCTACAGC 3030
 QY 141 ThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160

DB 3031 ACCCTTGGGGGTATTTTGGCTTCAACAGATTCCACTGCGCACTTTTCCACCACGCTAGCTGG 3090
 QY 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
 DB 3091 CAAGAGACTCATCAACAACAACTGGGGATTCCGACCCCAAGAGACTCAACTTCAAGCTCTTT 3150
 QY 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeu 200
 DB 3151 AACATTCAAGTCAAGAGGTACGAGAAATGACGGTACGACAGCAGATTGCCAATAACCTT 3210
 QY 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
 DB 3211 ACCAGCAGCTTTCAGGTGTTTACTGACTCGGAGTACCACTCCCGTACGCTCTCGCTCG 3270
 QY 221 AlaHisGlnGlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGly 240
 DB 3271 GCGCATCAAGGATGCTCCCGCGTTCCAGCAGAGCGTCTTATGGTGCACAGTATGGA 3330
 QY 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
 DB 3331 TACCTCACCTGAAACACGCGGAGTACGAGCAGTAGGACGCTCTTCACTTACTGCTGGAG 3390
 QY 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
 DB 3391 TACTTTCTTCTCAGATGCTGCTACCGGAAACAACCTTTACCTTACGCTACACCTTTTGG 3450
 QY 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
 DB 3451 GACGTTCTTTCACAGCAGCTACGCTCAACGCGAGTCTGGACGCTCTCATGAATCTT 3510
 QY 301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
 DB 3511 CTCATCGACCACTACCTGTTTACTTGGACAGACAAACACTCCCAAGTGGNACCACCG 3570
 QY 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
 DB 3571 CAGTCAAGGCTTCAGTTTTCTCAGCGCGGAGGAGTGACATTCGGGACCACTAGGAAC 3630
 QY 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
 DB 3631 TGGCTTCTTGGACCTTGTACCGCCAGCAGGAGTATCAAGACATCTCGGGATAACAAC 3690
 QY 361 AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 380
 DB 3691 AACAGTGAATACCTCGTGAGTGGAGTACCAAGTACCACCTCAATGGCAGAGACTCTCTG 3750
 QY 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet 400
 DB 3751 GTGAATCCGGGGCCGCCCATGGCAAGCCACAGGACGATGAAGAAAAGTTTTTCTCTCAG 3810
 QY 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
 DB 3811 AGCGGGGTCTCATCTTTGGGAAGCAAGGCTCAGAGAAAACAANTGTGAACATTGAAGAG 3870
 QY 421 ValMetIleThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
 DB 3871 GTCATGATTACACGACGAGGAAATCGGAACAACCAATCCCGTGGCTACGGAGCAGTAT 3930
 QY 441 GlyThrValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHis 460
 DB 3931 GGTTCTGTATCTACCAACCTCCAGAGAGGCAACAGACAGCAGCTCCCGCAGATGCAAC 3990
 QY 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro 480
 DB 3991 ACAAGAGCGCTTCTCCAGGCATGGTCTGGCAGACAGAGATGTGTACCTTTCAGGGGCC 4050
 QY 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
 DB 4051 ATCTGGGCAAGATTCCACACGCGGACATTTTCCACCTCTCCCTCATGGGTGGA 4110
 QY 501 PheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsn 520
 DB 4111 TTGGGACTTAAACACCCCTCTCCACAGATTCTCATCAAGAACACCCCGGTACCTGCGAAT 4170

QY 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
Db |||||
4171 CCTTCGACACCTTCAGTGGCGCAAGTTGCTTCTTCATCACACAGTACTCCACGGGA 4230
QY 541 Gln-ValSerValGluIleGluTTPGluLeuGlnLysGluAsnSerLysArgTTPAsnPr 560
Db |||||
4231 CAGGTCAGCGTGGAGATCGAGTGGAGCTGCAGAAAGGAAACAGCAACCTGGAATCC 4290
QY 560 oGluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAs 580
Db |||||
4291 CGAATTCAGTACACTTCCAACTACAAGTCTGTTAATCGTGGACTTACCGTGGATAC 4350
QY 580 nAsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db |||||
4351 TAATGCGTGTATTTCAGAGCTCGCCCAATTGGCCACGATACCTGACTCGTAATCTG 4408

RESULT 6

US-08-475-391-1
; Sequence 1, Application US/08475391
; Patent No. 5786211
; GENERAL INFORMATION:
; APPLICANT: Johnson, Philip R.
; TITLE OF INVENTION: Adeno-Associated Virus Materials and
; TITLE OF INVENTION: Methods
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,391
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/254,358
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5786211and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31975
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4680 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-475-391-1

Alignment Scores:
Pred. No.: 1.44e-253 Length: 4680
Score: 2722.50 Matches: 492
Percent Similarity: 89.00% Conservative: 42
Best Local Similarity: 82.00% Mismatches: 64
Query Match: 83.74% Indels: 2
DB: 1 Gaps: 1

US-09-807-802A-15 (1-599) x US-08-475-391-1 (1-4680)

QY 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSer 20

Db |||||
2614 ACGGCTCCGGGAAAAAGAGCGCGGTAGACACTCTCTCTGTGGAGCCAGACTCTCTCTCG 2673
QY 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGly 40
Db |||||
2674 GGAACCGAAGGCGGCGGCGAGAGCTGCAAGAAAAGATTGAATTTGGTCAGACTGGA 2733
QY 41 AsnSerGluSerValProAspProGlnProLeuGlyGluProProAlaIleThrProAlaAla 60
Db |||||
2734 GAGCGACTCAGTACCTGACCCCGAGCTCTCGGACAGCCAGCCAGAGCCCCCTCTGGT 2793
QY 61 ValGlyProThrThrMetAlaSerGlyGlyValAlaProMetAlaAspAsnAsnGluGly 80
Db |||||
2794 CTGGAACTAATACGATGCTACAGGAGTGGCGCCCAATGGCAGACAATTAACGAGGCG 2853
QY 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
Db |||||
2854 GCCGAGCGAGTGGGTAACTCTCCGGAATTTGGCATTGGCATTTCCACATGATGGGCGAC 2913
QY 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyr 120
Db |||||
2914 AGAGTCATCACACAGCAGCCCGAACCCTGGGCGCTGCCACCTACAAACAACCACTCTAC 2973
QY 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer 140
Db |||||
2974 AAACAAATTTCCAGCCCAATCA---GGAGCCTCGAACGACAATCACTACTTTTGGCTACAGC 3030
QY 141 ThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrp 160
Db |||||
3031 ACCCTTGGGGGTATTTGACTTCAACAGATTCCACTGCCACTTTTCCACCAGCTGACTGG 3090
QY 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
Db |||||
3091 CAAGACTCATCAACAACAATCGGGATTCCGACCCCAAGAGACTCAACTTCAAGCTCTTT 3150
QY 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeu 200
Db |||||
3151 AACATTCAAGTCAAGAGGTCAAGCAGAAATGACGTACGACGAGATTGCCAATTAACCTT 3210
QY 201 ThrSerThrValGlnValPheSerAspSerGlyTyrGlnLeuProTyrValLeuGlySer 220
Db |||||
3211 ACCAGCAGCTTCCAGGTGTTTACTGACTCGGAGTACCAGCTCCGCTACGCTCTCGGCTCG 3270
QY 221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGly 240
Db |||||
3271 GCGCATCAAGGATGCTTCCCGCGCTTCCAGCAGACGCTCTCATGGTGCCACAGTAGGA 3330
QY 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
Db |||||
3331 TACCTCACCTGAAACAACGGGAGTCAGGAGTAGGACGCTCTTCATTTTACTGCTGGAG 3390
QY 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
Db |||||
3391 TACTTCTCTCTCAGATGCTGCTACCGGAAACAACACTTTTACCTTCAGCTACACTTTTGG 3450
QY 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
Db |||||
3451 GAGCTTCTCTTCCACAGCAGCTACGCTCACAGCCAGAGTCTGGACCGCTCTCATGAATCCT 3510
QY 301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
Db |||||
3511 CTATCGACCACTACCTGTTATTTGAGCAGAAACAACACTCCAACTGGAGACCCACAGC 3570
QY 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
Db |||||
3571 CAGTCAAGGCTTCAGTCTTCTCAGCGCGGAGCGAGTGACATTCCGGGACCACTAGGAAC 3630
QY 341 TrpLeuProGlyProCysTyrArgGlnArgValSerLysThrLysThrAspAsnAsn 360
Db |||||
3631 TGGCTTCTCGACCTGTTTACCAGCAGCAGAGTATCAAAAGACATCTGCGGATAAACAC 3690
QY 361 AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 380
Db |||||

Db 3691 AACAGTGAATCTCTGGAGCTGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTG 3750
Qy 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspGluAspLysPhePheProMet 400
Db 3751 GTGAATCCGGGCGCCGATGCAAGGACGATGAAGAAAGTCTTTCTCTCAG 3810
Qy 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
Db 3811 AGCGGGTCTCATCTTTGGGAAGCAAGCTCAGAGAAACAAATGTGAACATTTGAAAG 3870
Qy 421 ValMetIleThrAspGluGluLysAlaThrAsnProValAlaThrGluArgPhe 440
Db 3871 GTCATGATTACAGCAGAGGAAATCGGAACCAATCCGCTGGCTAGCGAGCAGTAT 3930
Qy 441 GlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHis 460
Db 3931 GGTTCCTGTATCTACCAACCTCCAGAGAGGCAACAGACAGCAGTACCAGCATCTCAAC 3990
Qy 461 AlaMetGlyAlaLeuProGlyMetValTTPGlnAspArgAspValTyrLeuGlnGlyPro 480
Db 3991 ACACAGGCGTCTTCCAGGCGATGGTCTGGCAGGACAGAGATGTGTACCTTCAGGGGCC 4050
Qy 481 IleTTPAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
Db 4051 ATCTGGCAAGATTCCACACAGGAGCGGACATTTTCAACCTCTCCCTCATGGTGA 4110
Qy 501 PheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsn 520
Db 4111 TTGGAGCTTAAACACCCCTCTCCACAGATTCTCATCAAGAACACCCCGGTACTCGAAT 4170
Qy 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
Db 4171 CCTTCGACCACTTCAGTGGCGGAAGTTGTCTCTTCTCATCACAGTACTCCACGGA 4230
Qy 541 Gln-ValSerValGluIleGluTTPGluLeuGlnLysGluAsnSerLysArgTTPAsnPr 560
Db 4231 CACGGTCAGCTGGAGATCGAGTGGGAGCTGCAGAGAGGAAACAGCAACGCTGGAATCC 4290
Qy 560 OGUValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAs 580
Db 4291 CGAAATTCAGTACACTTCCAACCTACCAACAGTCTGTTAATCGTGGACTTACCGTGATAC 4350
Qy 580 nAnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db 4351 TAATGGGTGTATCAGAGCTCGCCCATTCGACCAAGATCTGACTGCTGTAATCTG 4408

RESULT 7

US-08-709-609-1
; Sequence 1, Application US/08709609
; Patent No. 5858775
; GENERAL INFORMATION:
; APPLICANT: Johnson, Philip R.
; TITLE OF INVENTION: Adeno-Associated Virus Materials and
; TITLE OF INVENTION: Methods
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/709,609
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5858775and, Greta E.

; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31975
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4680 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-709-609-1
Alignment Scores:
Pred. No.: 1,44e-253 Length: 4680
Score: 2722.50 Matches: 492
Percent Similarity: 89.00% Conservative: 42
Best Local Similarity: 82.00% Mismatches: 64
Query Match: 83.74% Indels: 2
Gaps: 1

US-09-807-802A-15 (1-599) x US-08-709-609-1 (1-4680).

Qy 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSer 20
Db 2614 ACAGCTCCGGGAAAAGAGGCGGTAGAGCACTCTCTCTGGAGCCAGACTCTCTCTCG 2673
Qy 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGly 40
Db 2674 GGAACCGGAAAGCGGCGCAGCAGCTGCAAGAAAAGATTGAATTTTGGTCAGACTGA 2733
Qy 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
Db 2734 GACGAGACTCAGTACTGACCCCGCCTCTCGACAGCACCAGCAGCCCCCTCTCGT 2793
Qy 61 ValGlyProThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGluGly 80
Db 2794 CTGGGAACATAATACGTGGCTACAGGAGTGGCGCACCACCAATGGCAGACAATAACGAGGC 2853
Qy 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTTPHisCysAspSerThrTTPLeuGlyAsp 100
Db 2854 GCGCAGCGAGTGGGTAAATCTCCGGAATTTGGCATTCGATTCACATGGATGGGCGCAG 2913
Qy 101 ArgValIleThrThrSerThrArgThrAlaLeuProThrTyrAsnAsnHisLeuTyr 120
Db 2914 AGAGTCATCACACCAGCAGCCGAACTGGGCGCTCCGCCACCTACACACCACTCTTAC 2973
Qy 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer 140
Db 2974 AAACAAATTTCCAGCCAAATCA--GGAGCCTCGAAGCACAATCACTACTTTGGCTACAGC 3030
Qy 141 ThrProTTPGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTTP 160
Db 3031 ACCCTTGGGGGTATTTTGACTTCAACAGATTCCACTGCGCATTTTTCACCACTGACTGG 3090
Qy 161 GlnArgLeuIleAsnAsnTTPGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
Db 3091 CAAGACTCATCAACAACTCTGGGATTCGCCCAAGAGACTCACTTCAAGTCTCTT 3150
Qy 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeu 200
Db 3151 AACATTCAAGTCAAGAGGTACGCGAATGACGGTACGAGCAGCATTTGCCAATAACCTT 3210
Qy 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
Db 3211 ACCAGCAGGTTTTCAGGTGTTTACTGACTCGGAGTACCACTCCCGTACGCTCTCGGCTCG 3270
Qy 221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGly 240
Db 3271 GCGCATCAAGGATGCTCTCCCGCTTCCAGCAGAGCGTCTTTCATGGTGCACAGTATGA 3330

```
QY 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
Db 3331 TACCTCACCCCTGAACAACGGAGTACAGGAGTACAGGAGTCTCTCTATTTACTGCTCGGAG 3390
QY 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
Db 3391 TACTTTCCTTCTCAGATGCTGGTACCGAACAACACTTTTACCTTCAGCTACACTTTTGG 3450
QY 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
Db 3451 GAGCTTCTCTTCCACAGCAGCTACGCTCACAGCCAGAGTCTGGACCGTCTCATGAATCT 3510
QY 301 LeuLeuAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
Db 3511 CTCATCGACCACTAGCTGATTTACTTGGAGCAGACAACACACTCCCAAGTGGACCAAC 3570
QY 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
Db 3571 CAGTCAGGCTTCAGTTTCTCAGCCGAGGAGTGCATTCGGGACCACTGCTAGGAAC 3630
QY 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
Db 3631 TGGCTTCTCGGACCTGTTTACCGCAGCAGCAGGATCAAGACATCTCGCGATTAACAAC 3690
QY 361 AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlnArgGluSerIle 380
Db 3691 AACAGTGAATACTCGTGGTACTGAGCTACCAAGTACCACTCAATGGCAGAGACTCTCG 3750
QY 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet 400
Db 3751 GTGAATCCGGGCGCCGATGCAAGCCACAGCAGATGAAGAAAGTTTTTCTCTAG 3810
QY 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
Db 3811 AGCGGGTTCATCTTTGGAGCAAGCTCAGAGAAACAAATGTGAACATTGAAGAAG 3870
QY 421 ValMetIleThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
Db 3871 GTCATGATTACAGACGAAGAAATTCGGAACAACCAATCTCCGCTGCTACGAGCAGTAT 3930
QY 441 GlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHis 460
Db 3931 GTTCTGTATCTACCACTTCAGAGGCAACAGACAGCAGCTACCGCAGATGTCAAC 3990
QY 461 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTrpLeuGlnGlyPro 480
Db 3991 ACACAGGCGCTTCTCCAGGCTGCTGCGAGGACAGAGATGTGTACCTTCAGGGGCC 4050
QY 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
Db 4051 ATCTGGGCAAGATTCACACAGCGGACGACATTTTCACCCCTCTCCCTCATGGTGA 4110
QY 501 PheGlyLeuLysAsnProProGlnIleLeuLysAsnThrProValProAlaAsn 520
Db 4111 TTCGGACTTAACACCTCTCTCCAGATCTCATCAAGAACACCCCGTACTCGGAT 4170
QY 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
Db 4171 CCTTCGACCACTTACGTGCGGCAAGTTTGTCTTCTTATCATACAGTACTCCAGGGA 4230
QY 541 Gln-ValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnPr 560
Db 4231 CACGGTCAGCGTGGAGATCGAGTGGAGCTGCAGAAAGGAAACAGCAACCGTGGATCC 4290
QY 560 oGluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAs 580
Db 4291 CGAAATTCAGTACACTTCCACTACCAAGTCTGTATTCGTGGACTTACCGTGGATAC 4350
QY 580 nAsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db 4351 TAATGGCGTATTACAGAGCTCGCCCAATGGCACCAGATACCTGACTCGTAATCTG 4408
```

RESULT 8

```
PCT-US95-07178-1
; Sequence 1, Application PC/TUS9507178
; GENERAL INFORMATION:
; APPLICANT: Johnson, Philip R.
; TITLE OF INVENTION: Adeno-Associated Virus Materials and
; TITLE OF INVENTION: Methods
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07178
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Noland, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31975
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4680 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US95-07178-1

Alignment Scores:
Pred. No.: 1,44e-253 Length: 4680
Score: 2722.50 Matches: 492
Percent Similarity: 89.00% Conservative: 42
Best Local Similarity: 82.00% Mismatches: 64
Query Match: 83.74% Indels: 2
DB: 5 Gaps: 1

US-09-807-802A-15 (1-599) x PCT-US95-07178-1 (1-4680)
QY 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSerSer 20
Db 2614 ACGGCTCGGGAAAAAGAGCGCGGTAGAGCACTCTCTGTGGAGCCAGACTCTCTCTCG 2673
QY 21 GlyIleGlyLysThrGlyGlnGluProAlaLysLysArgLeuAsnPheGlyGlnThrGly 40
Db 2674 GGAACCGAAAGGCGGCGCAGCGCTCGAAGAAAAAGATTGAATTTGGTCAGACTGA 2733
QY 41 AspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAlaAla 60
Db 2734 GACGAGACTCAGTACTACTGACCCCGAGCTCTCGAGACGCCACAGAGCCCTCTGCT 2793
QY 61 ValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGly 80
Db 2794 CTGGGAACCTAATACCATCGCTACAGGCAGTGGCGCAACCAATGGCAGACAATTAAC 2853
QY 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
Db 2854 GCCGACGAGTGGGTAATTCCTCCGAAATTTGGCATTTGCATTCACATGGATGGCGAC 2913
QY 101 ArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyr 120
Db 2914 AGAGTCATCACCAACCAAGCAGCAGCCGACCTGGGCGCTTGGCCACCTACACACCACT 2973
```

```
QY 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSer 140
Db 2974 AAACAAATTTCCAGCCAAATCA---GGAGCCTCGAAGCAGCAATCACTACTTTGGCTACAGC 3030
QY 141 ThrProTTPGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTTP 160
Db 3031 ACCCTTTGGGGTATTTTGACTTCAACAGATTCACCTGCCACTTTTTCACACAGTGACTGG 3090
QY 161 GlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPhe 180
Db 3091 CAAGACTCATCAACAACAACTGGGGATTCGACCCCAAGAGACTCAACTTCAAGCTCTTT 3150
QY 181 AsnIleGlnValLysGluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeu 200
Db 3151 AACATTCAAGTCAAGAGGTCCGAGAGTACGGTACGACGACATTTGCCAATAACCTT 3210
QY 201 ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 220
Db 3211 ACCAGCAGGTTCAGGTGTTTACTGACTCGGAGTACCACTCCCTAGCTCTCGGCTCG 3270
QY 221 AlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyrGly 240
Db 3271 GGCATCAAGGATGCTCTCCGCGCTTCCAGCAGACGCTTTCATGGTGCACAGTATGGA 3330
QY 241 TyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlu 260
Db 3331 TACCTCACCTCGAACAACGGGAGTCAGGAGTAGGACGCTCTTCATTTTACTGCTGGAG 3390
QY 261 TyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu 280
Db 3391 TACTTTCTCTTCAGATGCTGGTACCGGAAACAACTTTACCTTCAGCTACACTTTTGAG 3450
QY 281 GluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnPro 300
Db 3451 GACGTTCTCTTCCACAGCAGCTACGCTCACAGCCAGAGTCTGGACGCTCTCATGAATCT 3510
QY 301 LeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGln 320
Db 3511 CTCATCGACGAGTACCTGTATTACTTTGACGACCAACAAACACTCCCAAGTGGAAACACG 3570
QY 321 AsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsn 340
Db 3571 CAGTCAAGGCTTCAGTTTCTCAGCGCGGAGGAGTGACATTCGGGACCACTAGGAAC 3630
QY 341 TrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsn 360
Db 3631 TGCGTTCTCTGGACCTGTATTACCGCCAGCAGCGAGTATCAAGAGCATCTCGCGATAACAAC 3690
QY 361 AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 380
Db 3691 AACAGTGAATACCTCGTGGAGTGGAGTACCAAGTACCACTCAATGGCAGAGACTCTCG 3750
QY 381 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMet 400
Db 3751 GTGAATCCGGGGCCCGCCATGGCAAGCCCAAGGACGATGAAGAAAAGTTTTTCTCTCAG 3810
QY 401 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 420
Db 3811 AGCGGGGTTCTCATCTTTGGGAAGCAAGGCTCAGAGAAAACAATGTGAACATTGAAAG 3870
QY 421 ValMetIleThrAspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 440
Db 3871 GTCATGATTCACAGCAGAGAGGAATTCGGAACCAACCAATCCCTGGCTACGGAGCAGTAT 3930
QY 441 GlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHis 460
Db 3931 GGTTCTGTATCTACCACTCCAGAGAGGCAACAGACAGCAGTACCGCAGATGTCAAC 3990
QY 461 AlaMetGlyAlaLeuProGlyMetValTTPGlnAspArgAspValTyrLeuGlnGlyPro 480
Db 3991 ACACAAGGCGTTCTTCCAGGCATGGTCTGCGCAGGACAGAGATGTGTACCTTCAGGGGCC 4050
```

```
QY 481 IleTrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGly 500
Db 4051 ATCTGGCAAAAGATTCCACACGGACGGACATTTTTCACCCCTCTCCCTCATGGGTGA 4110
QY 501 PheGlyLeuLysAsnProProGlnIleLeuLysAsnThrProValProAlaAsn 520
Db 4111 TTGGACTTAAACACCCCTCTCTCCACAGATTCTCATCAAGAACACCCCGGTACCTGCGAAT 4170
QY 521 ProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGly 540
Db 4171 CCTTCGACCACTTCAGTGGGCAAGTTTGTCTTTCATCACACAGTACTCCACGGGA 4230
QY 541 GlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTTPAsnPr 560
Db 4231 CACGCTCAGCGTGGAGATCGAGTGGAGCTGCAGAGGAAAACAGCAACGCTGGAATCC 4290
QY 560 oGluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAs 580
Db 4291 CGAAATTCAGTACACTTCCAACTACAAAGTCTGTTAATCGTGGACTTACCGTGGATAC 4350
QY 580 nAsnGlyLeuTyrThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db 4351 TAATGGCGTGTATTTCAGAGCCTCGCCCAATTGGCACCAGATACCTGACTGTAATCTG 4408

RESULT 9
US-09-438-268-2
; Sequence 2, Application US/09438268
; Patent No. 6491907
; GENERAL INFORMATION:
; APPLICANT: Rabinowitz, Joseph E.
; APPLICANT: Samulski, Richard J
; APPLICANT: Xiao, Weidong
; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING
; FILE OF INVENTION: THE SAME
; FILE REFERENCE: 5470-186
; CURRENT APPLICATION NUMBER: US/09/438,268
; CURRENT FILING DATE: 1999-11-10
; EARLIER APPLICATION NUMBER: 60/107,840
; EARLIER FILING DATE: 1998-11-10
; EARLIER APPLICATION NUMBER: 60/123,651
; EARLIER FILING DATE: 1999-03-10
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; TYPE: DNA
; ORGANISM: Virus
US-09-438-268-2

Alignment Scores:
Pred. No.: 9,23e-222 Length: 8151
Score: 2398.50 Matches: 435
Percent Similarity: 82.17% Conservative: 58
Best Local Similarity: 72.50% Mismatches: 96
Query Match: 73.78% Indels: 11
DB: 4 Gaps: 3

US-09-807-802A-15 (1-599) x US-09-438-268-2 (1-8151)
QY 1 ThrAlaProGlyLysGlyArgProValGluGlnSerProGlnGluProAspSerSer 20
Db 2544 ACGGCTCTCGAAAGAGAGACCGTGTGATTGAATCCCCCAGCAGCCCGACTCTCCACG 2603
QY 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysArgLeuAsnPhe---GlyGlnThr 39
Db 2604 GGTATCGGCAAAAAGAGCAGCGCGCTAAAAGAGACTCGTTTTCGAAGACGAAACT 2663
QY 40 GlyAspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAla 59
Db 2664 GGAGCAGGCGGACGGACCCCTCGAGGATCAACTTCCGGA-----GCC 2705
QY 60 AlaValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGlu 79
Db 60 AlaValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGlu 79
```


LOCATION: 750
OTHER INFORMATION: n = g, a, c or t (u)
NAME/KEY: misc feature
OTHER INFORMATION: AAV4 caspid protein VPI gene
US-09-532-594B-5

Alignment Scores:

| | | | |
|------------------------|-----------|---------------|------|
| Pred. No.: | 1,396-167 | Length: | 2208 |
| Score: | 1830.50 | Matches: | 355 |
| Percent Similarity: | 69.87% | Conservative: | 74 |
| Best Local Similarity: | 57.82% | Mismatches: | 154 |
| Query Match: | 56.31% | Indels: | 31 |
| DB: | 4 | Gaps: | 10 |

US-09-807-802A-15 (1-599) x US-09-532-594B-5 (1-2208)

```
Qy 1 ThrAlaProGlyValLeuArgProValGluGlnSerProGlnGluProAspSerSerSer 20
Db 409 ACGGCTCTCTGGAAGAAGAGACCGGTGATTGAATCCCCCAGCAGCCGCGACTCTCCACG 468
Qy 21 GlyIleGlyLeuThrGlyGlnGlnProAlaLysLysArgLeuAsnPhe---GlyGlnThr 39
Db 469 GGTATCGGCAAAAAGGCAAGCAGCGCGCTAAAGAAGACTCGTTTCGAAGACGAACACT 528
Qy 40 GlyAspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAla 59
Db 529 GGAGCAGCGCGGACCCCTCGAGGATCAACTTCCGGA-----GCC 570
Qy 60 AlaValGlyProThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGlu 79
Db 571 ATGCTGTGATGACAGTGAGATGCGTCAGCAGCGTGGCGAGCTGTCAGTCGAGGGGGGACAA 630
Qy 80 GlyAlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTripleuGly 99
Db 631 GGTGCGGATGGAGTGGTAATACCTCGGTGATGGCATTCGCATTCCACCTGGTCTCAG 690
Qy 100 AspArgValIleThrSerThrArgThrTrpAlaLeuProThrTyAsnAsnHisLeu 119
Db 691 GCCCAGCTCAGCACCAGCAGCAGCAACCTGGTCTTGCCCACTCAACAACCAACCCCTN 750
Qy 120 TyrLysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyr 139
Db 751 TACAAGCGACTCGGAGAG-----AGCTGCGAGTCCAAACCTTACACCGGATTC 798
Qy 140 SerThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAsp 159
Db 799 TCCACCCCTCGGGATACCTTGACTTCAACCGCTTCCACTGCGCATCTCTCACCACGTCAC 858
Qy 160 TrpGlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeu 179
Db 859 TGGCAGCGACTCATCAACAACAACTGGGGCATGCGCAACCCAAAGCCATCGGGTCAAAATC 918
Qy 180 PheAsnIleGlnValIleGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsn 199
Db 919 TTCAACATCCAGTCAAGGAGTCAAGCATCGTCAACGCGGAGAGCAACCGTGGCTTAATAC 978
Qy 200 LeuThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGly 219
Db 979 CTTACCAGCAGCGTTTCAGATCTTGGGACTCGTCTGACGAACCTGCCGTACGTGATGAT 1038
Qy 220 SerAlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIleProGlnTyr 239
Db 1039 CGGGGTCAAGAGGCGGACCTGCTCTCTTTCCCAACAGCGTCTTATGTTGCGCCAGTAC 1098
Qy 240 GlyTyr-----LeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPhe 256
Db 1099 GGCTACTGTGGACTGTGGTACCGGCAACACTTCGCGACACAGACTGACAGAAATGCCCTTC 1158
Qy 257 TyrCysLeuGluTyrPheProSerGlnMetLeuAtgThrGlyYAsnAsnPheThrPheSer 276
Db 1159 TACTGCTGAGTACTTCTCTTCGACAGATGCTCGGAGCTGGCAACAACACTTCAAAATTAGC 1218
Qy 277 TyrThrPheGluGluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArg 296
```

```
Db 1219 TACAGTTTTCGAAAGGTGCTCTTCCACTCGATGTACGCGCACAGCCAGAGCTCGGACCGG 1278
Qy 297 LeuMetAsnProLeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsn----- 314
Db 1279 CTGATGAACCTCTCTCATCGACCACTGCTGTGGGACTGCATCGACACACACCGGAAC 1338
Qy 315 -----GlnSerGlySerAlaGlnAsnLysAspLeuLeuPheSerArgGlySerProAla 332
Db 1339 ACCCTGAATCCGGGACTGCCACCAAC-----TTTACCAAGCTCGGCGCTACC 1389
Qy 333 GlyMetSerValGlnProLysAsnTrpLeuProGlyProCysTyrArgGlnGlnArgVal 352
Db 1390 AACTTTTCCAACTTTTAAAGAACTGCTGCTCGCGGCTTCAATCAAGCAGCAGCGGCTTC 1449
Qy 353 SerLysThrLysThrAspAsnAsnAsnSerAsnPheThrTrpThrGlyAlaSer----- 370
Db 1450 TCAAGACTGCC-----AATCAAACTACAGATCTCCGCCCGGCTCAGACAGTCTC 1503
Qy 371 ---LysTyr-----AsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr 385
Db 1504 ATCAAAATACGAGCGCACAGCACTCTCGACGGAAGATGGAGTGCCTTGACCCCGGACCT 1563
Qy 386 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 405
Db 1564 CCAATGGCCACGCTGGAGCTCGGACCTCGGACGACGAAG---TTCAGCAACACGCCAGCTCATCTTT 1620
Qy 406 PheGlyLysGluSerAlaGlyAlaSerAsnTrpAlaLeuAspAsnValMetIleThrAsp 425
Db 1621 CGCGGGCTAAACAGAACGCGCAACACCGCCACCTGCTCCGGGACTCTGATCTTCCACTCT 1680
Qy 426 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 445
Db 1681 GAGGAGGAGCTGGCAGCCACCAACCGCACGATACGACGATGTGGGCAACCTACCTGGC 1740
Qy 446 AsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 465
Db 1741 GGTGACGAGACCAACAGCAACCTGCTCGGACCGTGACAGACTGACAGCTTGGGAGCGGTG 1800
Qy 466 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 485
Db 1801 CCTGANTGTCTGGCAAAACAGAGACATTTACTACAGGGTCCCATTTGGGCCAAGATT 1860
Qy 486 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 505
Db 1861 CCTCATCCGATGACACTTTCACCCCTCACCGCTGATTTGGTGGTGGTGGCTGGAACAC 1920
Qy 506 ProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 525
Db 1921 CCGCTCTCTCAAAATTTTATCAAGAACACCCCGGTACTCTGCAATCTCTGCAACGACCTTC 1980
Qy 526 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 545
Db 1981 AGCTTACTCGGTAAACTCTCTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACT 2040
Qy 546 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrThr 565
Db 2041 ATTGCTGGAGATCCAGAGAGCGGTTCAAACGCTGGAACCCCGAGGTCCAGTTTACC 2100
Qy 566 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 585
Db 2101 TCCAACCTACGACAGCAAAACTCTCTGTGTGGGCTCCCGATGCGGTGGGGAATACACT 2160
Qy 586 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db 2161 GAGCTAGGGCTATCGGTACCGCTACTCTACCCACCACTG 2202
```

RESULT 12

US-09-532-594B-1
; Sequence 1, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.

```
/ APPLICANT: Kotin, Robert M.
/ APPLICANT: Safer, Brian
/ APPLICANT: Davidson, Beverly
/ TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
/ FILE REFERENCE: 14014.025202
/ CURRENT APPLICATION NUMBER: US/09/532,594B
/ CURRENT FILING DATE: 2000-03-22
/ NUMBER OF SEQ ID NOS: 22
/ SOFTWARE: FASTSEQ for Windows Version 4.0
/ SEQ ID NO 1
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
/ OTHER INFORMATION: synthetic construct
/ NAME/KEY: misc feature
/ LOCATION: 3009_
/ OTHER INFORMATION: n = g, a, c or t(u)
/ NAME/KEY: misc feature
/ OTHER INFORMATION: AAV4 genome
/ US-09-532-594B-1
```

```
Alignment Scores:
Pred. No.: 4,75e-167 Length: 4767
Score: 1830.50 Matches: 355
Percent Similarity: 69.87% Conservative: 74
Best Local Similarity: 57.82% Mismatches: 154
Query Match: 56.31% Indels: 31
DB: 4 Gaps: 10
```

US-09-807-802a-15 (1-599) x US-09-532-594B-1 (1-4767)

```
QY 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnGluProAspSerSer 20
DB 2668 ACGGCTCTGGAAAGAGAGCGTTGATTGAATCCCGCCAGCGCGACTCTCCACG 2727
QY 21 GlyIleGlyThrGlyGlnGlnProAlaLysLysArgLysPhe---GlyGlnThr 39
DB 2728 GGTATCGGCAAAAGAGAGCGAGCGCGCTTAAAGAGAGCTCGTTTCAGACGCAAACT 2787
QY 40 GlyAspSerGluSerValProAspProGlnProLeuGluProProAlaThrProAla 59
DB 2788 GGAGCAGGCGAGCGGACCCCTGAGGATCACTTCGGA-----GCC 2829
QY 60 AlaValGlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGlu 79
DB 2830 ATGCTGTGATGATGATGATGCGTGCAGCAGCTGCGCGAGCTGCAGTCGAGGGGGGACAA 2889
QY 80 GlyAlaAspGlyValGlyAsnAlaSerGlyAsnThrHisCysAspSerThrTrpLeuGly 99
DB 2890 GGTGCCGATGGATGGGTAAATGCTCGGGTGATGGCATTTGCCATTCACCTGGTCTGAG 2949
QY 100 AspArgValIleThrThrSerThrArgThrTrpAlaLeuProThrTrpAsnAsnHisLeu 119
DB 2950 GGCACGTCAGCACCAACAGCAGCAGACCTGGGTCTTGCCCACTTACACCAACCACTTN 3009
QY 120 TyrLysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyr 139
DB 3010 TACAAGCGACTCGGAGAG-----AGCCTGCAGTCCCAACCACTTACAACGGATTC 3057
QY 140 SerThrProTrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAsp 159
DB 3058 TCCACCCCTCGGGGATATTGTACTTCAACCGCTTCCACTGCGCCACTTCTCACCACGTGAC 3117
QY 160 TrpGlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeu 179
DB 3118 TGGCAGGCACTCATCAACACCACTGGGGCATGGACCCCAAGCCATCGCGGTCAAAATC 3177
QY 180 PheAsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsn 199
DB 3178 TTCAACATCAGGTCAAGGAGGTCAACGACGTGCAACCGGAGACCAACCGTGGCTAATAAC 3237
```

```
QY 200 LeuThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGly 219
DB 3238 CTTACCAGCAGCGTTTCAGATCTTTGCGGACTCGTGTACGAACCTGCCGTGATGAT 3297
QY 220 SerAlaHisGlnGlyCysLeuProPheProAlaAspValPheMetIleProGlnTyr 239
DB 3298 GCGGGTCAAGAGGCGCGCTGCTCTTTCCCAACGAGCGTCTTTATGGTCCCGCAGTAC 3357
QY 240 GlyTyr-----LeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPhe 256
DB 3358 GGCTACTGTGACTGGTGCAGCGCAACACTTCGCGACCAACAGACTGCAGAAATGCCTTC 3417
QY 257 TyrCysLeuGluTyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSer 276
DB 3418 TACTGCTTGGAGTACTTTCCTTCAGATGCTGGGACTGCGCAACACTTTGAAATACG 3477
QY 277 TyrThrPheGluGluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArg 296
DB 3478 TACAGTTTGGAGAGTGGCTTTCCTCTCGATGTACGCGCACAGCAGCAGAGCTCGACGG 3537
QY 297 LeuMetAsnProLeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsn----- 314
DB 3538 CTGATGAACCTCTCATCGACCACTGCTGGGGACTGCAATCGACCAACCGGAAACC 3597
QY 315 -----GlnSerGlySerAlaGlnAsnLysAspLeuLeuPheSerArgGlySerProAla 332
DB 3598 ACCCTGAATGCGGAGCTGCCACCAAC-----TTTACCAAGCTCGGCGCTACC 3648
QY 333 GlyMetSerValGlnProLysAsnTrpLeuProGlyProCysTyrArgGlnGlnArgVal 352
DB 3649 AACTTTTCCAACTTTAAAGAACTGGCTGCCCGCGCTTCAATCAACGACGAGCGCTTC 3708
QY 353 SerLysThrLysThrAspAsnAsnAsnSerAsnPheThrTrpThrGlyAlaSer----- 370
DB 3709 TCNAAGACTGCC-----AATCAAAACTACAAGATCCCTGCCCGGCTCAGACAGTCTC 3762
QY 371 ---LysTyr-----AsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr 385
DB 3763 ATCAATACGAGACGACGACGACCTCTGGACGAGATGAGTGCCTGACCCCGGAGCT 3822
QY 386 AlaMetAlaSerHisLysAspGluAspLysPhePheProMetSerGlyValMetIle 405
DB 3823 CCNATGGCCAGCGCTGGAGCTCGGACGACGAGCAAG---TTCAGCAACGACGAGCTCATCTT 3879
QY 406 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 425
DB 3880 GCGGGGCTTAACAGAACGCAACGCGCCACCGTACCCGGGACTCTGATCTTCACTCT 3939
QY 426 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 445
DB 3940 GAGGAGGAGCTGGCAGCCACCAACCGCAGATACGAGATGTGGGCAACCTACCTGGC 3999
QY 446 AsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 465
DB 4000 GGTGACGAGCAACAGCAACCTGCGGACCGTGGGACAGACTGCAGCCTTGGGAGCGGTG 4059
QY 466 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaValIle 485
DB 4060 CTGGAATGGTCTGCAAAACAGACATTTACTACCGGGTCCCTATTTGGGCCAAGATT 4119
QY 486 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 505
DB 4120 CCTCATACGATGGACACTTTTCCACCTTCCCGCTGATGGTGGTGGTGGCTGAAACAC 4179
QY 506 ProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 525
DB 4180 CGGCTCTCTCAAAATTTTATCAAGAACACCCCGGTACTCGCAATCTCTGCAACAGCTTTC 4239
QY 526 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 545
DB 4240 AGCTTACTCCCGGTAACTCTTCAATTTACTAGTACAGACCTGCGCAGGTGTCGGTGCAG 4299
QY 546 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrThr 565
```

Db 4300 ATTGACTGGAGATCCAGAGCGGTCCAAAGCTGGAACCCGAGGTCAGTTTACC 4359
Qy 566 SerAenTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAenGlyLeuTyrThr 585
Db 4360 TCCAACTACGACAGCAAACTCTCTGTGTGGGCTCCCGATGCGGTGGGAAATACACT 4419
Qy 586 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
Db 4420 GAGCTAGGCTATCGGTACCGCTACCTACCCACCACCTG 4461

RESULT 13

US-09-438-268-1/c
; Sequence 1, Application US/09438268
; Patent No. 6491907
; GENERAL INFORMATION:
; APPLICANT: Rabinowitz, Joseph E.
; APPLICANT: Samulski, Richard J.
; APPLICANT: Xiao, Weidong
; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 5470-186
; CURRENT APPLICATION NUMBER: US/09/438,268
; CURRENT FILING DATE: 1999-11-10
; EARLIER APPLICATION NUMBER: 60/107,840
; EARLIER FILING DATE: 1998-11-10
; EARLIER APPLICATION NUMBER: 60/123,651
; EARLIER FILING DATE: 1999-03-10
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 1
; LENGTH: 7214
; TYPE: DNA
; ORGANISM: Virus
US-09-438-268-1

Alignment Scores:

Pred. No.: 9,21e-167 Length: 7214
Score: 1830.50 Matches: 355
Percent Similarity: 69.87% Conservative: 74
Best Local Similarity: 57.82% Mismatches: 154
Query Match: 56.31% Indels: 31
Gaps: 10

US-09-807-802A-15 (1-599) x US-09-438-268-1 (1-7214)

Qy 1 ThrAlaProGlyLysValArgProValGluGlnSerProGlnGluProAspSerSerSer 20
Db 2575 ACGGCTCTCTGGAAAGAGAGACCGTTGATTGAATCCCCCAGCGCCGACTCTCTCCACG 2516
Qy 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPhe---GlyGlnThr 39
Db 2515 GGTATCGGCAAAAGGCAAGCGCGCTAAAGAGAGCTGTTTTCGAAGACGAACT 2456
Qy 40 GlyAspSerGluSerValProAspProGlnProLeuGlyGluProProAlaThrProAla 59
Db 2455 CGAGCAGCGCAGCGACCCCTGAGGATCAACTTCCGGA-----GCC 2414
Qy 60 AlaValGlyProThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAnGlu 79
Db 2413 ATGCTGTATGACAGATGGATGGTGCAGCGTGGCGGAGTGCAGTGGGGGGGAGCA 2354
Qy 80 GlyAlaAspGlyValGlyAsnAlaSerGlyAsnThrHisCysAspSerThrTrpLeugly 99
Db 2353 GGTGCGGATGAGTGGGTAAATGCTCGGTGATTGGCATTCGATTCCACCTGGTCTGAG 2294
Qy 100 AspArgValIleThrThrSerThrThrThrTrpAlaLeuProThrTyrAsnAsnHisLeu 119
Db 2293 GGCCACGTCACGACCAGCAGCAGACCTGGGTCTTGGCCACCTACCAACCAACCCCTC 2234
Qy 120 TyrLysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyr 139
Db 2233 TACAAGCGACTCGGAGAG-----AGCCTGCAGTCCAACTACCAACGGATTTC 2186

Qy 140 SerThrProTyrGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAsp 159
Db 2185 TCCACCCCTGGGATACTTTGACTTCAACCGCTTCCACTGCCACTTCTCACCACGCTGAC 2126
Qy 160 TrpGlnArgLeuIleAsnAsnAsnTyrPheArgProLysArgLeuAsnPheLysLeu 179
Db 2125 TGGCAGCGACTCATCAACAACAATCGGGGATCGACCCCAAGCCATCGCGGTCAANAATC 2066
Qy 180 PheAsnIleGlnValLysGluValThrThrAsnAspGlyValThrThrIleAlaAsnAsn 199
Db 2065 TTCAACATCCAGGTCAAGGAGGTCAAGCTCGAACCGCGAGACAGCGGTGGCTTAATAC 2006
Qy 200 LeuThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGly 219
Db 2005 CTTACCAGCAGCGTTTCAGATCTTTGCGACTCGTCGTACGAACCTGCGCTGATGATGAT 1946
Qy 220 SerAlaHisGlnGlyCysLeuProPheProPheProAlaAspValPheMetIleProGlnTyr 239
Db 1945 GCGGGTCAAGAGGGGCGACCTGCTCTTTTCCCAACGACGTCTTTATGGTGCCCGACGATC 1886
Qy 240 GlyTyr-----LeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerSerPhe 256
Db 1885 GGCTACTGTGAGCTGGTGACCGGCAACACTTCGCGAGCAACAGACTGCAGACAAATGCTTTC 1826
Qy 257 TyrCysLeuGluTyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSer 276
Db 1825 TACTGCTGGAGTACTTCTCTCGCAGATGCTCGGAGCTGCAACAACATTTGAAATATACG 1766
Qy 277 TyrThrPheGluGluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArg 296
Db 1765 TACAGTTTTCGAGAAGGTGCTTCTCCACTCGATGTACGCGCACAGCCAGAGCTCGACCGG 1706
Qy 297 LeuMetAsnProLeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsn----- 314
Db 1705 CTGATGAACCTCTCATCGACAGTACTCTGGGAGCTGCAATCGACCAACCACCGAACC 1646
Qy 315 -----GlnSerGlySerAlaGlnAsnLysAspLeuLeuPheSerArgGlySerProAla 332
Db 1645 ACCCTGAATCGCGGAGCTGCCACCACAC-----TTTACCAAGCTCGCGCTTACC 1595
Qy 333 GlyMetSerValGlnProLysAsnTyrLeuProGlyProCysTyrArgGlnGlnArgVal 352
Db 1594 AACTTTTCCAACTTTAAAGAACTGCTGCTCGCGGCTTCAATCAAGCAGCAGGCTTTC 1535
Qy 353 SerLysThrLysThrAspAsnAsnAsnSerAsnPheThrTrpThrGlyAlaSer----- 370
Db 1534 TCANAGACTGCC-----ATCAAAACTACAAAGATCCTGCCACCGGTCAGACAGTCTC 1481
Qy 371 ---LysTyr-----AsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr 385
Db 1480 ATCAAAATACAGACGACAGCACTCTGACGGAAGATGGAGTGCCTGACCCCGGACCT 1421
Qy 386 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 405
Db 1420 CCAATGGCCAGCTGGTGGAGCTGCGGACGAGCAAG---TTCAGCAACAGCAGCTCATCTTT 1364
Qy 406 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 425
Db 1363 GCGGGGCTAAACAGACGACGCAACCGCCACCTGACCCGGGACTGTGATCTTACCTCT 1304
Qy 426 GluGluLysLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 445
Db 1303 GAGGAGGAGCTGGCAGCCACCAACCGCCATACGAGCATGTGGGGCACTTACCTGGC 1244
Qy 446 AsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 465
Db 1243 GGTCAACAGACAAACAGCAACCTTCGCGAGCAGACTGCAGCAGCTTGGAGCCGTG 1184
Qy 466 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 485
Db 1183 CTTGGATGTCTGGCAAAACAGAGACATTTACTACCGGGTCCCATTTGGGCCAAGATT 1124

QY 486 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuIleAsn 505
DB 1123 CCTCATACCGATGGACACTTTTCCACCCCTCACCGCTGATTGGTGGTTTGGGCTGAACAC 1064
QY 506 ProProProGlnIleLeuIleLeuAsnThrProValProAlaAsnProProAlaGluPhe 525
DB 1063 CGCGCTCTCAATTTTATCAGAACACCCCGGCTACTCGCATCTCGCAACGACCTTC 1004
QY 526 SerAlaThrLysPheAlaSerPheIleThrGlnIleThrSerThrGlyGlnValSerValGlu 545
DB 1003 AGCTTACTCCGGTAAATCTCTTATTACTCAGTACAGCACTGGCCAGGTGCGTGCAG 944
QY 546 IleGluThrGluLeuGlnIleGluAsnSerLysArgTProAsnProGluValGlnThr 565
DB 943 ATTGACTGGGAGATCCAGAGAGCGGTCCAAACGCTGGAACCCCGAGGTCCAGTTTACC 884
QY 566 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 585
DB 883 TCCAACTACGGACACAAACTCTCTGTTGGGCTCCCGATCGCGCTGGGAATACACT 824
QY 586 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 599
DB 823 GAGCCTAGGGCTATCGGTACCGCTACCTCACCCACCACTG 782

RESULT 14

US-09-532-594B-19
; Sequence 19, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.
; APPLICANT: Kotin, Robert M.
; APPLICANT: Davidson, Beverly
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.025202
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 1617
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; OTHER INFORMATION: synthetic construct
; NAME/KEY: misc_feature
; LOCATION: 162
; OTHER INFORMATION: n = g, a, c or t(u)
; NAME/KEY: misc_feature
; OTHER INFORMATION: AAV4 capsid protein VP3 gene
US-09-532-594B-19

Alignment Scores:
Pred. No.: 3,19e-154 Length: 1617
Score: 1690.50 Matches: 325
Percent Similarity: 71.30% Conservative: 65
Best Local Similarity: 59.41% Mismatches: 132
Query Match: 52.00% Indels: 25
DB: 4 Gaps: 9

US-09-807-802a-15 (1-599) x US-09-532-594B-19 (1-1617)

QY 67 AlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGlyAsn 86
DB 7 GCAGCAGCTGGCGAGCTGCAGTCTC---GAGGSGGACAAAGGTGCGCATGGAGTGGGTAAT 63
QY 87 AlaSerGlyAsnThrHisCysAspSerThrTyrLeuGlyAspArgValIleThrThrSer 106
DB 64 GCCTCGGGGTATGGGATTCGATTCACCTGGTCTTGAGGGGCCACCTCAGCACCAACAC 123
QY 107 ThrArgThrTyrAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerAla 126

DB 124 ACCAGAACCTGGGCTCTTGCCCACTACAACAACACCCTTACAAGCGACTCGGAGAG--- 180
QY 127 SerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTyrGlyTyrPhe 146
DB 181 -----AGCTGCGAGTCCAAACACCTACAACGGATTCTCCACCCCTCGGGATACCTTT 231
QY 147 AspPheAsnArgPheHisCysHisPheSerProArgAspTyrGlnArgLeuIleAsnAsn 166
DB 232 GACTTCAACCGCTTCACTGCCACTTCTCAACACGCTGACTGGCAGCGACTCATCAACAC 291
QY 167 AsnTyrGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLysGlu 186
DB 292 AACTGGGGCATCGAACCCAAAGCATGCGGGTCAAAATCTTCAACATCCAGGTCAAGGAG 351
QY 187 ValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGlnVal 206
DB 352 GTCACGACGTCGACACGGCGAGACACGGTGGCTTAATTAACCTTACCACGCGTTCAGATC 411
QY 207 PheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCysLeu 226
DB 412 TTTCGGGACTCGTCGTACGAACTGCCGTAGTGTAGTGGGTCAAGAGGGCAGCTG 471
QY 227 ProProPheProAlaAspValPheMetIleProGlnTyrGlyTyr-----LeuThr 243
DB 472 CCTCCTTTTCCCAACGAGCTCTTTATGGTGGTCCCGTACGCTACTGTGGACTGGTGACC 531
QY 244 LeuAsnAsnGlySerGlnIleValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 263
DB 532 GGCAACACACTTTCGACGACACGACTGACAGAAATGCCTTCTACTGCTGGAGTACTTTCCT 591
QY 264 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 283
DB 592 TCGCAGATGCTGCGGACTGGCAACACTTTGAAATTAGGTACAGTTTGGAGAGTGCT 651
QY 284 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 303
DB 652 TTCCACTCGATGTACGGGCACAGCAGCGCTGACCGGCTGTATGAACCCCTCTCATCGAC 711
QY 304 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsn-----GlnSerGlySerAla 319
DB 712 CAGTACCTGTGGGACTGCAATCGACACCAACCGAACCCCTGTAATGCCGGGACTGCC 771
QY 320 GlnAsnLysAspLeuLeuPheSerArgLysSerProAlaGlyMetSerValGlnProLys 339
DB 772 ACCACCAAC-----TTTACCAAGCTGGCGCTTACCAACTTTTCCAACTTTAAAAAG 822
QY 340 AsnTyrLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsn 359
DB 823 AACTGGCTGCCCGGGCTTCAATCAAGCAGCAGCGGCTTCTCAAAAGACTGCC-----AAT 876
QY 360 AsnAsnSerAsnPheThrTyrThrGlyAlaSer-----LysTyr----- 372
DB 877 CAAAACTACAAGATCCCTGCCCGCGGTGCAGACAGTCTCATCAATACGACGACGACGAC 936
QY 373 AsnLeuAsnGlyArgGluSerIleAsnProGlyThrAlaMetAlaSerHisLysAsp 392
DB 937 ACTCTGGACGGAAGATGGAGTGCCCTGACCCCGGACCTCCCAATGGCCACGCTGGACCT 996
QY 393 AspGluAspLysPhePheProMetSerGlyValMetIlePheGlyLysGluSerAlaGly 412
DB 997 GCGGACAGCAAG---TTCAAGCAACAGCCAGCTCATCTTTGCGGGGCTTAAACAGAACGCG 1053
QY 413 AlaSerAsnThrAlaLeuAspAsnValMetIleThrAspGluGluIleLysAlaThr 432
DB 1054 AACACGGCCACCGTACCCGGGACTCTGATCTTCACTCTGAGGGAGAGCTGGCAGCCACC 1113
QY 433 AsnProValAlaThrGluArgPheGlyThrValAlaValAsnPheGlnSerSerThr 452
DB 1114 AACGCCACCGATACGGACATGTGGGGCAACCTACCTGGCGGTGACACGACGACCAACAGCA 1173
QY 453 AspProAlaThrGlyAspValHisAlaMetGlyAlaLeuProGlyMetValTyrGlnAsp 472
DB 1174 CTGCGGACCGTGGACAGACTGACAGCCTTGGAGCCGCTGGGAATGGTCTGGCAAAAC 1233

QY 473 ArgAspValTyrLeuGlnGlyProIleThrAlaLysIleProHisThrAspGlyHisPhe 492
Db 1234 AGAGACATTACTACAGGGTCCCATTTGGCCAGAGATTCTCATACGATGGACATTT 1293
QY 493 HisProSerProLeuMetGlyGlyPheGlyLeuLysAsnProProGlnIleLeuIle 512
Db 1294 CACCCTCAGCGCTGATTGGTGGGTTTGGGCTGAACACACCGCGCTCTCTCAAAATTTTATC 1353
QY 513 LysAsnThrProValProAlaAsnProProAlaGluPheSerAlaThrLysPheAlaSer 532
Db 1354 AAGAACACCCCGGTACCTCGCAATCTCTGCAAGCCTTCAGCTCTACTCCGGTAAACTCC 1413
QY 533 PheIleThrGlnTyrSerThrGlyGlnValSerValGluIleGluTyrGluLeuGlnLys 552
Db 1414 TTCATTACTAGTACAGCACTGGCCAGGTGCGGTGCAGATTGACTGGGAGATCCAGAAG 1473
QY 553 GluAsnSerLysArgTyrAsnProGluValGlnTyrThrSerAsnTyrAlaLysSerAla 572
Db 1474 GAGCGGTCCAAACCGTGGAAACCCGAGGTCCAGTTTACCTCCAACACAGCAAAAC 1533
QY 573 AsnValAspPheThrValAspAsnAsnGlyLeuTyrThrGluProArgProIleGlyThr 592
Db 1534 TCTCTGTGTGGCTCCGATCGGCTGGGAATACACTGAGCCTAGGGCTATCGGTACC 1593
QY 593 ArgTyrLeuThrArgProLeu 599
Db 1594 CGTACTCACCCACCACCTG 1614

RESULT 15

US-09-438-268-3
; Sequence 3, Application US/09438268
; Patent No. 6491907
; GENERAL INFORMATION:
; APPLICANT: Rabinowitz, Joseph E.
; APPLICANT: Samulewski, Richard J.
; APPLICANT: Xiao, Weidong
; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 5470-186
; CURRENT APPLICATION NUMBER: US/09/438,268
; CURRENT FILING DATE: 1999-11-10
; EARLIER APPLICATION NUMBER: 60/107,840
; EARLIER FILING DATE: 1998-11-10
; EARLIER APPLICATION NUMBER: 60/123,651
; EARLIER FILING DATE: 1999-03-10
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2271
; TYPE: DNA
; ORGANISM: Virus
US-09-438-268-3

Alignment Scores:
Pred. No.: 6,528-65 Length: 2271
Score: 772.00 Matches: 195
Percent Similarity: 47.55% Conservative: 87
Best Local Similarity: 32.88% Mismatches: 257
Query Match: 23.75% Indels: 54
Db: 4 Gaps: 16

US-09-807-802A-15 (1-599) x US-09-438-268-3 (1-2271)

QY 1 ThrAlaProGlyLysLysArgProValGluGlnSerProGlnProAspSerSer 20
Db 412 ACGGCTCCGGAAAGAGGCGGTAGAGCACTCTCTGTGGAGCCAGACTCTCTCTCG 471
QY 21 GlyIleGlyLysThrGlyGlnGlnProAlaLysLysArgLeuAsnPheGlyGlnThrGly 40
Db 472 GGAACCGGAAGCGCGGCCAGCTGCAAGAAAAGATTGAATTTTGGTCAGACTGA 531
QY 41 AspSerGluSerValProAspProGlnProLeuGlyGluProAlaThrProAlaAla 60

Db 532 GAGCAGACTCAGTACTGACCCAGCCTCTCGGACAGCACCCAGCAGCCCCCTCTGGT 591
QY 61 ValGlyProThrThrMetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGly 80
Db 592 CTGGGAACATAATACGATGACTTCAGTTAAATTCGCAGAA--GCCAGCAGCTGGTGCAGGA 648
QY 81 AlaAspGlyValGlyAsnAlaSerGlyAsnTyrHisCysAspSerThrTyrLeuGlyAsp 100
Db 649 GGGGGGGCAGTAATCTGTCAAAAGCATGTCGAGTGGAGTGGGGGCCACTTTTAGTGTAAAC 708
QY 101 ArgValIleThrThrSerThrArgThrTyrAlaLeuProThrTyrAsnAsnHisLeuTyr 120
Db 709 TCTGTAACTTGTACATTTTCCAGACAGTCTTTTAAATTCATATGACCCAGACACCATAT 768
QY 121 LysGlnIleSerSerAlaSerThrGlyAlaSerAsnAsp----- 133
Db 769 AAGTGTCTTCTCCGAGCGAGTAGCTGCCACATGCCAGTGGAAAGGAGGCAAGGTT 828
QY 134 -----AsnHisTyrPheGlyTyrSerThrProTyrPheAspPheAsnArg 150
Db 829 TGCACCATCAGTCCCATTAATGGGATCTCAACCCATGGAGATATTTAGATTTTAAATGCT 888
QY 151 PheHisCysHisPheSerProArgAspTyrGlnArgLeuIleAsnAsnThrPheGlyPhe 170
Db 889 TTAATTTATTTTTCACCTTTAGAGTTTTCAGCAGCTTAATTGAATAATTATGGAAGTATA 948
QY 171 ArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLysGluValThr----- 188
Db 949 GCTCTCTGATGCTTTAACTGTAACCATATCAGAAATTCCTGTTAAGGATGTTACAGACAA 1008
QY 189 ThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGlnValPheSer 208
Db 1009 ACTGGAGGGGGGTA---CAGGTTACTGCAGCAGCTACAGCGGCTATGCATGTTAGTA 1065
QY 209 AspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCysLeuProPro 228
Db 1066 GACCATGAATACAGTACCCATATGTTAGGGCAAGTCAGGATCTTTAGCCCCCAGAA 1125
QY 229 PheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsnAsn---Gly 247
Db 1126 CTTCCTATTTGGGTATACTTTCCCTCAATATGCTTACTTACAGTAGGAGATGTTAAC 1185
QY 248 SerGlnAlaValGly-----ArgSerSerPheTyrCys 258
Db 1186 ACACAAGGAATTTCTGGAGACAGCAAAATTAGCAAGTGAAGAAATCAGCATTTTATGTT 1245
QY 259 LeuGluTyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThr 278
Db 1246 TTGGAACACAGTTCTTTTTCAGCTTTTAGGTACAGGAGGTACAGCACTATGCTCTTATAAG 1305
QY 279 PheGluGluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMet 298
Db 1306 TTTCTCTCAGTGGCCCCCAGAAATTTAGAGGGCTGCGAGTCAACACTTTTATGAATGTAC 1365
QY 299 AsnProLeuIleAspGlnTyrLeuTyrLeuAsnArgThrGlnAsnGlnSerGlySer 318
Db 1366 AATCCCTTA-----TACGGATCCCGCTTAGGGTCTCTGACACATTAGGAGGTGAC 1416
QY 319 AlaGlnAsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnPro 338
Db 1417 CCAAAATTTAGATCTTTA-----ACACATGAAGACCATGCAATTCAGGCC 1461
QY 339 LysAsnThrLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAsp 358
Db 1462 CAAAACTTCATGCGAGGCCCATCTAGTAACTCAGTGTCTACAAAGGAGGAGACAGCTCT 1521
QY 359 AsnAsnAsnSerAsnPheThrTyrThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGlu 378
Db 1522 AATAGTGGAGTGAAGAGCCCTTAACAGGCTTAGCAGAGGTACCTCTCAAAACACATAGA 1581
QY 379 SerIleLeuAsnProGly---ThrAlaMetAlaSerHisLysAspAspGluAspLysPhe 397

Db 1582 ATATCTTACGCCCTGGCCAGTGTCTCAGCCATACCACCTGGACACAGATAAATAT 1641
QY 398 PheProMetSerGlyValMetIlePheGlySerGlySerAlaGlyAlaSerAsnThrAla 417
Db 1642 GTCACAGGAATAATGCCATTCTCATGTGCAGACCACTATGGT-----AACGCT 1692
QY 418 LeuAspAsnVal-----MetIleThrAspGluGluIleLys 430
Db 1693 GAAGACAAGAGATATCAGCAAGAGTGGGTAGATTTCCTCAATGAAAAAGACAGCTAAAA 1752
QY 431 AlaThrAsnProValAlaThrGluArgPheGlyThrValAlaValAsnPheGlnSerSer 450
Db 1753 CAGTTACAGGGTTTAAACATGCACACCTAC-----TTTCCCAATAAA 1794
QY 451 SerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeuProGlyMetValTrp 470
Db 1795 GGAACCCAGCAATATACAGATCAAAATTGAG---CGCCCCCTAATGGTGGGTCTGTATGG 1851
QY 471 GlnAspArgAspValTyrLeuGlnGlyProIleTyrAlaValIleProHisThrAspGly 490
Db 1852 AACAGAAGAGCCCTTCACATGAAAGCCAGCTGTGGAGTAAATTCCTCAATTTAGATGAC 1911
QY 491 HisPheHisProSer---ProLeuMetGlyGlyPheGlyLeuLysAsnProProGln 509
Db 1912 AGTTTAAAACTCAGTTTCAGCCTTAGGAGGATGGGGTTTGCATCAGCCACCTCCTCAA 1971
QY 510 IleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPheSerAlaThrLys 529
Db 1972 ATATTTTAAAA-----ATATTACCACAAAGTGGGCCCAATTGGAGGTATTAAATCAATG 2025
QY 530 PheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGluIleGluTrpGlu 549
Db 2026 GGAATTACTACCTTAGTTCAGTATGCCGTGGGAATTATGACAGTAACATGACATTTAAA 2085
QY 550 Leu---GlnLysGluAsnSerLysArgTyrAsnProGlu 561
Db 2086 TTGGGGCCCGTAAAGCTACGGGACGGTGGAAATCCTCAA 2124

Search completed: February 15, 2004, 01:37:39
Job time : 249.456 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 14, 2004, 15:26:50 ; Search time 3324.15 seconds
(without alignments)
4379.586 Million cell updates/sec

Title: US-09-807-802A-15
Perfect score: 3251
Sequence: 1 TAPGKRPEVQSPQBPDS...NNGLYTERPIGTRYLTRPL 599

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-Q=cn2_1/USPTO.spool/US09807802/runat_11022004_175608_15941/app query.fasta_1.2389
-DB=EST -OPMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODES=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09807802 @CN_1_1 7257 @runat_11022004_175608_15941 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_etc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hcc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_nam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|----------|--------------------|
| 1 | 140.5 | 4.3 | 2318 | 11 | BC014681 | BC014681 Homo sapi |
| 2 | 134 | 4.1 | 856 | 28 | BH164736 | BH164736 ENTTS28TF |
| 3 | 125.5 | 3.9 | 2146 | 11 | AK042727 | AK042727 Mus muscu |
| 4 | 125.5 | 3.9 | 2295 | 11 | AK035953 | AK035953 Mus muscu |
| 5 | 125 | 3.8 | 754 | 10 | BG207226 | BG207226 RST26694 |
| 6 | 121.5 | 3.7 | 925 | 28 | BH152154 | BH152154 ENTPV12TR |
| 7 | 115.5 | 3.6 | 753 | 28 | BH115587 | BH115587 RPCI-24-3 |
| 8 | 114.5 | 3.5 | 1042 | 13 | BQ642977 | BQ642977 AGENCOURT |
| 9 | 114 | 3.5 | 5809 | 11 | AK076994 | AK076994 Mus muscu |
| 10 | 113.5 | 3.5 | 2637 | 11 | BC037597 | BC037597 Mus muscu |
| 11 | 113.5 | 3.5 | 2688 | 11 | BC019128 | BC019128 Mus muscu |
| 12 | 113 | 3.5 | 1960 | 11 | AY103595 | AY103595 Zea mays |
| 13 | 113 | 3.5 | 4782 | 11 | AK048546 | AK048546 Mus muscu |
| 14 | 112.5 | 3.5 | 1651 | 11 | AY104980 | AY104980 Zea mays |
| 15 | 112 | 3.4 | 3057 | 11 | AK044947 | AK044947 Mus muscu |
| 16 | 111.5 | 3.4 | 3329 | 11 | AK033012 | AK033012 Mus muscu |
| 17 | 111.5 | 3.4 | 4124 | 11 | AK031710 | AK031710 Mus muscu |
| 18 | 110.5 | 3.4 | 768 | 14 | CD558053 | CD558053 AGENCOURT |
| 19 | 110.5 | 3.4 | 815 | 10 | BG746848 | BG746848 602704187 |
| 20 | 110.5 | 3.4 | 925 | 13 | BQ440075 | BQ440075 AGENCOURT |
| 21 | 110.5 | 3.4 | 2904 | 11 | AK046602 | AK046602 Mus muscu |
| 22 | 110.5 | 3.4 | 3619 | 11 | AK028748 | AK028748 Mus muscu |
| 23 | 110.5 | 3.4 | 4843 | 11 | AK082447 | AK082447 Mus muscu |
| 24 | 110 | 3.4 | 640 | 14 | CA356072 | CA356072 628112 NC |
| 25 | 110 | 3.4 | 701 | 14 | CA353789 | CA353789 638563 NC |
| 26 | 110 | 3.4 | 703 | 14 | CA378088 | CA378088 658613 NC |
| 27 | 110 | 3.4 | 902 | 29 | CNS03388 | AL225809 Tetraodon |
| 28 | 109.5 | 3.4 | 2365 | 11 | AK081012 | AK081012 Mus muscu |
| 29 | 109 | 3.4 | 350 | 9 | AW545004 | AW545004 C0188E06- |
| 30 | 109 | 3.4 | 581 | 13 | BQ163384 | BQ163384 952077G03 |
| 31 | 109 | 3.4 | 659 | 13 | BW314066 | BW314066 BW314066 |
| 32 | 109 | 3.4 | 687 | 13 | BU723756 | BU723756 SJWAZH04 |
| 33 | 108.5 | 3.3 | 838 | 10 | BE642450 | BE642450 Cr12_6_B0 |
| 34 | 108 | 3.3 | 497 | 12 | BM417785 | BM417785 952005F11 |
| 35 | 108 | 3.3 | 2125 | 11 | AK079113 | AK079113 Mus muscu |
| 36 | 108 | 3.3 | 2931 | 11 | AK019448 | AK019448 Mus muscu |
| 37 | 108 | 3.3 | 4784 | 11 | AK041115 | AK041115 Mus muscu |
| 38 | 107 | 3.3 | 411 | 12 | BM152500 | BM152500 TCBAPE91 |
| 39 | 107 | 3.3 | 544 | 12 | BM896109 | BM896109 952063E01 |
| 40 | 107 | 3.3 | 550 | 29 | CC171347 | CC171347 i17id11.b |
| 41 | 107 | 3.3 | 576 | 12 | BM173665 | BM173665 952003C01 |
| 42 | 107 | 3.3 | 582 | 12 | BM417581 | BM417581 952018G12 |
| 43 | 107 | 3.3 | 582 | 13 | BU049515 | BU049515 111109D0 |
| 44 | 107 | 3.3 | 588 | 12 | BM895805 | BM895805 952065E09 |
| 45 | 107 | 3.3 | 672 | 12 | BM333577 | BM333577 MEST157-G |

ALIGNMENTS

RESULT 1
BC014681
LOCUS BC014681 2318 bp mRNA linear HTC 04-OCT-2001
DEFINITION Homo sapiens, Similar to osa, clone IMAGE:3866722, mRNA.
ACCESSION BC014681
VERSION BC014681.1 GI:15928395
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2318)
AUTHORS Strausberg,R.

| | |
|---------|---|
| TITLE | Direct Submission |
| JOURNAL | Submitted (01-OCT-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA |
| REMARK | NIH-MGC Project URL: http://mgc.nci.nih.gov |
| COMMENT | Contact: MGC help desk Email: cgabs-r@mail.nih.gov Tissue Procurement: ATCC DNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/ Contact: villalon@bcm.tmc.edu Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A. |

| | | | |
|----|------|---|------|
| Qy | 94 | ApsSerThrTripleLeuGlyAspArgValIleThrThrSerThrArgThrTrpAlaLeuPro | 113 |
| Db | 1054 | -----GGAAATCCTTTTATTGCCACCTCAGGA-----CCT | 1083 |
| Qy | 114 | ThrTrpAsnAsnHisLeuTyrlsYsGlnIleSerSerAlaserThrGlyAlaSerAsnAsp | 133 |
| Db | 1084 | GCCACTTCGCACGTGCCCGCCAGCAGAGTCCCAGCATGCACACCTTCCTTGCGTCACTCG | 1143 |
| Qy | 134 | -----AsnHisTyrrPheGlyTyrrSerThrProTyrGlyTyrPheAppPheAsn | 149 |
| Db | 1144 | GTGCAGCAGTTCATCACCAACC-CCTCTACTGCTCTCCATGG----- | 1184 |
| Qy | 150 | ArgPheHisCys-His-----PheSerProArgAspTrpGlnArgLeuIleAsnAs | 166 |
| Db | 1185 | AGNATCCGTTGCCACAGTCCCGAGATTCTCCCCGAATCCTCCCAACA----- | 1233 |
| Qy | 166 | nAenTrpGlyPheArgProGlyArgLeuAenPheLysLeuPheAsnIleGlnVallysGl | 186 |
| Db | 1234 | ----GGGGTGTGTAGCGCCCAAACCTTAACCTTTAGTTCTCGGAGCGACAGTCCCTC | 1289 |
| Qy | 186 | uValThrThrAsnAspGlyValThrThrIleAlaAsnAenLeuThr-SerThrValGlnVa | 206 |
| Db | 1290 | TCTCTACTATAAACAC-----ProPheProAlaAs | 232 |
| Qy | 206 | lPheAspSerGluTyrrGlnLeuProTyrrValLeuGlySerAlaHisGlnGlyCysLe | 226 |
| Db | 1306 | ---TCAGGCGCAGTATTCTCGATATCCTTAC-----AGTAACCTTAATCAGGGATTAGT | 1355 |
| Qy | 226 | u-----ProPheProAlaAs | 232 |
| Db | 1356 | TACANATACAGGAGTGAATCAAAATTTAGCGCTTACANAATAACTCCAATGAATCACTC | 1415 |
| Qy | 232 | pValPheMetIleproGlnTyrglyTyrrLeuThrLeuAsnAsnGlySerGlnAlaValGl | 252 |
| Db | 1416 | CGTA-----CCAAGATAC-----CCCAATGCTGTAGG | 1442 |
| Qy | 252 | yArgSerSerPheTyrrCysLeuGluTyrrPheProSerGlnMetLeuArgThrGlyAsnAs | 272 |
| Db | 1443 | A-----ATGCACACAGCAGCCCATCCACCCAGTGGCTCACTTAACCA | 1511 |
| Qy | 272 | nPheThrPheSerTyrrPheGluGluValProPheHisSerSerTyrrAlaHisSerGl | 292 |
| Db | 1467 | ACTA-----ATGCACACAGCAGCCCATCCACCCAGTGGCTCACTTAACCA | 1511 |
| Qy | 292 | nSerLeuAspArgLeuMetAsnProIleIleAspGlnTyrrLeuTyrrTyrrLeuAsnArgTh | 312 |
| Db | 1512 | AATGAACACACAACATATGATCCT-----TCACGAGCC | 1544 |
| Qy | 312 | rGlnAenGlnSerGlySerAlaGlnAenLysAspLeuLeuPheSerArgGlySerProAl | 332 |
| Db | 1545 | TCAGGNACTTATGCTCTCCACCTCCATGTCACCCATGNAAGCATGAGTATCCAGC | 1604 |
| Qy | 332 | aGlyMetSerValGlnProLysAenTrpLeuProGlyProCysTyrrArgGlnGlnArgva | 352 |
| Db | 1605 | AGGC-----ACTCCTCCTCCACAAGTCAGCGCCGGGAAG-TGCTGG----- | 1643 |
| Qy | 352 | lSerLysThrLysThrAspAenAsnAenSerAenPheThrTrpThrGlyAlaSerLysTy | 372 |
| Db | 1644 | -----GATACCAATGGAAATTTGGCAG-----TTATCCAAATA | 1675 |
| Qy | 372 | rAsnLeuAenGlyArgGluSerIleIleAenProGlyThrAlaMetAlaSerHisLysAs | 392 |
| Db | 1676 | T-----GCCCATCCTCAGCCATCTCACACGCCCTCGTGCCATGGGAATCGGAC | 1726 |
| Qy | 392 | pAspGluAspLysPhePheProMetSerGlyValMetIlePheGlyLysGluSerAlaGl | 412 |
| Db | 1727 | AGAGGA-----TATGGGCCCCAGAACATGTC | 1755 |
| Qy | 412 | yAlaSerAsnThrAlaLeuAspAenValMetIleThrAsp-----GluGluGl | 428 |
| Db | 1754 | AGCAGTCTGTCCTATTTATAGGCATGTCCTCGGCACCAAGGAATTTGATGGGCACATGA | 1813 |
| Qy | 428 | uileLys-----AlaThrAenProValAlaThrGl | 438 |

```

1814 GGCACAAAGTGTCTGCTGTTGGCTTGGAGACCCACAGCAATCCAGGACGACTGA 1873
Qy 438 uargPheGlyThrValAlaValAsnPhGlnSerSerSerThrAspProAlaThrGlyAs 458
Db 1874 TACCTGGCCACACACATCC-----TGGTCAACAGCCATCTTTTCAGCAGTTGCCAA 1924
Qy 458 pValHisAlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuG1 478
Db 1925 CCGTCTCTCCATCGAGCTCACCGGG-----CITGC 1957
Qy 478 nGlyProIleTrpAlaLysIle-ProHisThrAsp-----GlyHisPheHisProSerP 496
Db 1958 ACCACCACTCTTCACTCCACACCTCATCCAGCCTTGGGCACAGCTCCACCCATCAC 2017
Qy 496 roLeuMetGlyGlyPheGlyLeuLysAsnProProGlnIleLeuLysAsnThrP 516
Db 2018 CC-----CAGAACCCCGCAGAAAGTG-----C 2041
Qy 516 roValProAlaAsnProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrG 536
Db 2042 CTGTGCATCAGCATTCCTCGGTCGGAGCCCTTCTAGAGAACCCAGTGCGCGATGATGCTC 2101
Qy 536 InTyrSerThrGlyGlnValSerVal-----GluLeuGluT 548
Db 2102 AGGTTAGTGGACCGCAATGCTCAGCTAGTGTAGAGAGTGATGATTACCTGCCATCAATAGA 2161
Qy 548 rpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrThrSerAsnT 568
Db 2162 AGCAGCCACAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2221
Qy 568 yrAlaLysSerAlaAsnValAspPheThr-----ValAspAsnAsnGlyLeu 583
Db 2222 GTAAAGGTTTGGTAAAGATGACTCTCCCTGGTGGGTAGATTAACCAAGAACTA 2274

```

RESULT 2

```

BH164736
LOCUS      BH164736
DEFINITION ENTTS28TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, genomic survey sequence.
ACCESSION  BH164736
VERSION    BH164736.1 GI:15738174
KEYWORDS   GSS.
SOURCE     Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE  Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
            Determination of clone end sequences from Entamoeba histolytica
            HMI:IMSS sheared DNA library (2001)
            Unpublished
JOURNAL    Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@igr.org
            Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
            DNA library
Seq primer:
Class: shotgun
High quality sequence start: 17
High quality sequence stop: 629.
Location/Qualifiers

```

FEATURES

source

```

1..856
/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pHOS1; Site 1: Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.

```

Genomic DNA isolated from broth cultures of *E. histolytica* using a method described by Clark and Diamond (Clark, C.G. and Diamond, L.S. (1993) *Entamoeba histolytica*: a method for isolate identification. *Exp. Parasitol.* 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In *Genome Sequencing: A Practical Approach*, eds. M. Vaudin and B. Borell, Oxford University Press, 1999).

BASE COUNT 347 a 229 c 89 g 191 t
ORIGIN

Alignment Scores:

Pred. No.: 0.0811 Length: 856
Score: 134.00 Matches: 88
Percent Similarity: 33.42% Conservative: 44
Best Local Similarity: 22.28% Mismatches: 140
Query Match: 4.12% Indels: 123
DB: 28 Gaps: 19

US-09-807-802a-15 (1-599) x BH164736 (1-856)

```

Qy 160 TrpGlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPhLeuLeu 179
Db 9 TGGAGACACCTCATTTGGTAAAT-----TCTTCATTCGCTGGA 47
Qy 180 PheAsnIleGlnValLysGluValThrThrAsnAspGlyValThrThr----- 195
Db 48 TTTAATACTAACCAACACGTCGGGACAACTAATCCGTTCAATCTCTCTGGAAGTTCAAGTC 107
Qy 196 --IleAlaAsnAsnLeuThrSerThrValGlnValPheSerAspSerGluTyrGlnLeu 214
Db 108 TCAGGACAAATAACCCCTTTGCTACTACTCA-----AACATACACAACT 155
Qy 215 ProTyrValLeuGlySerAlaHisGlnGlyCysLeuProPheProAlaAspValPhe 234
Db 156 CTTTTCACACAGGACAACTAAC-----CCATTTAAT----- 188
Qy 235 MetIleProGlnTyrGlyTyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSer 254
Db 189 -----ACTACTAACCAATACTACTACTCA----- 212
Qy 255 SerPheTyrCysLeuGluTyrPheProSerGlnMetLeuArgThrGlyAsnAsnPhThr 274
Db 213 -----ACAATAATCCATTAAT 230
Qy 275 PheSerTyrThrPheGluGluValProPheHisSerSerTyrAlaHisSerGlnSerLeu 294
Db 231 ACTAATACTACAGCACCAACATCCATTTAATAACAAATACTACTACTAAT----- 278
Qy 295 AsparGluMetAsnProLeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsn 314
Db 279 -----AATCCATTT-----AGTACTAATAATAC 299
Qy 315 GlnSerGlySerAlaGlnAsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMet 334
Db 300 ACTACCAATGGAACCAACCAAT-----TTAATTAATCAA-----ACTCCTCTCTCTAAT 347
Qy 335 SerValGlnProLysAsnTrpLeuProGlyProCysTyrArgGlnGlnArgValSerLys 354
Db 348 ACTATAACACCCAGGAAT-----AATACCAACTACT 377
Qy 355 ThrLysThrAspAsnAsnAsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeu 374
Db 378 ACAACAGAGAGTAATCCATTTGGTAATTTCACTACCACTAAT----- 419
Qy 375 AsnGlyArgGluSerIleAsnProGlyThrAlaMetAlaSerHisLysAspAspGlu 394
Db 420 AATACTACCTCCAGTACACTTCAACAGGAACACCATCACACAGGAAGTAATCCATTT 479
Qy 395 AspLysPhePheProMetSerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSer 414

```

Genomic DNA isolated from broth cultures of *E. histolytica* using a method described by Clark and Diamond (Clark, C.G. and Diamond, L.S. (1993) *Entamoeba histolytica*: a method for isolate identification. *Exp. Parasitol.* 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In *Genome Sequencing: A Practical Approach*, eds. M. Vaudin and B. Borell, Oxford University Press, 1999).

BASE COUNT 347 a 229 c 89 g 191 t
ORIGIN

Alignment Scores:

Pred. No.: 0.0811 Length: 856
Score: 134.00 Matches: 88
Percent Similarity: 33.42% Conservative: 44
Best Local Similarity: 22.28% Mismatches: 140
Query Match: 4.12% Indels: 123
DB: 28 Gaps: 19

US-09-807-802a-15 (1-599) x BH164736 (1-856)

```

Qy 160 TrpGlnArgLeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPhLeuLeu 179
Db 9 TGGAGACACCTCATTTGGTAAAT-----TCTTCATTCGCTGGA 47
Qy 180 PheAsnIleGlnValLysGluValThrThrAsnAspGlyValThrThr----- 195
Db 48 TTTAATACTAACCAACACGTCGGGACAACTAATCCGTTCAATCTCTCTGGAAGTTCAAGTC 107
Qy 196 --IleAlaAsnAsnLeuThrSerThrValGlnValPheSerAspSerGluTyrGlnLeu 214
Db 108 TCAGGACAAATAACCCCTTTGCTACTACTCA-----AACATACACAACT 155
Qy 215 ProTyrValLeuGlySerAlaHisGlnGlyCysLeuProPheProAlaAspValPhe 234
Db 156 CTTTTCACACAGGACAACTAAC-----CCATTTAAT----- 188
Qy 235 MetIleProGlnTyrGlyTyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSer 254
Db 189 -----ACTACTAACCAATACTACTACTCA----- 212
Qy 255 SerPheTyrCysLeuGluTyrPheProSerGlnMetLeuArgThrGlyAsnAsnPhThr 274
Db 213 -----ACAATAATCCATTAAT 230
Qy 275 PheSerTyrThrPheGluGluValProPheHisSerSerTyrAlaHisSerGlnSerLeu 294
Db 231 ACTAATACTACAGCACCAACATCCATTTAATAACAAATACTACTACTAAT----- 278
Qy 295 AsparGluMetAsnProLeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsn 314
Db 279 -----AATCCATTT-----AGTACTAATAATAC 299
Qy 315 GlnSerGlySerAlaGlnAsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMet 334
Db 300 ACTACCAATGGAACCAACCAAT-----TTAATTAATCAA-----ACTCCTCTCTCTAAT 347
Qy 335 SerValGlnProLysAsnTrpLeuProGlyProCysTyrArgGlnGlnArgValSerLys 354
Db 348 ACTATAACACCCAGGAAT-----AATACCAACTACT 377
Qy 355 ThrLysThrAspAsnAsnAsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeu 374
Db 378 ACAACAGAGAGTAATCCATTTGGTAATTTCACTACCACTAAT----- 419
Qy 375 AsnGlyArgGluSerIleAsnProGlyThrAlaMetAlaSerHisLysAspAspGlu 394
Db 420 AATACTACCTCCAGTACACTTCAACAGGAACACCATCACACAGGAAGTAATCCATTT 479
Qy 395 AspLysPhePheProMetSerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSer 414

```

```

Db      480  GGTAAATTTCTACTCCCTCAAACTCAGCTCCCAACAGGAAATTAATCAACAGGAACAGCT 539
Qy      415  AenThrAlaLeuAspAenValMetIleThrAspGluGluGluIleYsAlaThrAenPro 434
Db      540  ACACACACAGGAAATTAATCAACATCACA-----ACAGGAAGTAATCCA 584
Qy      435  ValAlaThrGluArgPheGlyThrValAlaValAsnPheClnSerSerThrAspPro 454
Db      585  -----TTTGGT-----AATTTCACTCCCTCAACACCACTCCC 617
Qy      455  AlaThrGlyAspValHisAlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAsp 474
Db      618  ACCACAGGGGAAAT-----AATATCAGGAGCAACTACCAACAGAGAAAT 662
Qy      475  ValTyxLeuGlnGlyProIleTrpAlaLeuIleProHisThrAspGlyHisPheHisPro 494
Db      663  AAATACACAGGA-----ACAGGAAGTATCCATTT-----GGTAATTTCACTACT 707
Qy      495  SerProLeuMetGlyGlyPheGlyLeuLeuAsnProProProGlnIleLeuIle-LysAs 514
Db      708  TCAAAATCTACTACCCACACAGGAAACAACTACACAGGAAATTAACACACTACAGGAA 767
Qy      514  nThrProValProAlaAenProProAlaGluPheSerAlaThrLysPheAlaSerPhe1 534
Db      768  CAACCCACACCCGGGAAACCCCAACACCGGAAACCCCA-----810
Qy      534  eThrGlnTyxSerThrGlyGlnValSerValGluIleGluTrp 548
Db      811  -----ACAGGAAGGATCCTTTTGGAAATTTTCATAC 840

RESULT 3
AK042727
LOCUS
DEFINITION
  AK042727      2146 bp      mRNA      linear      HTC 05-DEC-2002
  Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length
  enriched library, clone:A730019105 product:hypothetical protein,
  full insert sequence.
ACCESSION
  AK042727.1 GI:26335284
KEYWORDS
  HTC; CAP trapper.
SOURCE
  Mus musculus (house mouse)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
  1
REFERENCE
  Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
  Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  Normalization and subtraction of cap-trapper-selected cDNAs to
  prepare full-length cDNA libraries for rapid discovery of new genes
  Genome Res. 10 (10), 1617-1630 (2000)
  20493374
  11042159
REFERENCE
  3
  Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
  Konno,H., Akiyama,J., Nishi,K., Kiteunai,T., Tashiro,H., Itoh,M.,
  Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
  Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
  Fujiwaki,S., Inoue,K., Togawa,Y., Iizawa,M., Ohara,E., Wataniki,M.,
  Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
  Okazaki,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
  Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  RIKEN integrated sequence analysis (RISA) system--384-format
  sequencing pipeline with 384 multicapillary sequencer
  Genome Res. 10 (11), 1757-1771 (2000)
  20530913
  11076861
REFERENCE
  4
  Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,

```

```

Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
Aizawa,K., Iizawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamataka,I.,
Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
Pleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
Kuehl,P., Lewis,S., Mateuio,Y., Nikaido,I., Pesole,G.,
Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
Wagner,J., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
Baldarrelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
Carninci,P., de Bonaldo,M.P., Brownstein,M.J., Bult,C.,
Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
Tayo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilmking,L.,
Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
and Hayashizaki,Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6921), 685-690 (2001)
21085660
11217851
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2146)
ADACHI,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tegami,M., Tagawa,A., Takahashi,F., Takata-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/.
FEATURES
    source
        1. .2146
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="C57BL/6J"
            /db_xref="FANTOM DB:A730019105"
            /db_xref="taxon:10090"
            /clone="A730019105"
            /tissue_type="cerebellum"
            /clone_lib="RIKEN full-length enriched mouse cDNA library"
            /dev_stage="7 days neonate"
            253_->2145
            /note="unmated protein product; hypothetical protein
            (evidence: Truncated-Longest-ORF)
            putative"
            /codon_start=1
    CDS

```


| | | | | | | | |
|----|------|---|----------------|-----------|-----------|-------|-----|
| Qy | 31 | LysLys----- | -----ArgLeuAsn | PheGlyGln | ThrGlyAsp | Ser | 42 |
| Db | 791 | CGAAAGCAGCGGCAATCTTTTGTGGATGCGGCTCCATGTGGGGCCCCGGGCCGTTC | ----- | ----- | ----- | ----- | 850 |
| Qy | 43 | GluservValProAspProGlnProLeuGlyuProProAlaThrProAlaAlaValGly | 62 | | | | |
| Db | 851 | GGTACAGACACAGATACGAGCCCCCTA-----CCAGCAGCAGCGCCAGCGCGC | 901 | | | | |
| Qy | 63 | ProThrThrMetAlaSerGlyGlyValAlaProMetAlaAspAsnAsnGluGlyAlaAsp | 82 | | | | |
| Db | 902 | ---TCCTTCGGGGCCCCGGCACAGGGCCACCCCCAGCA-----CATGCAGCAGAT | 949 | | | | |
| Qy | 83 | GlyValGlyAsnAlaSerGlyAsnTrpHis-----CysAspSerThrTrpLeuGly | 99 | | | | |
| Db | 950 | -----GGGCAGCTATCTGGCAGCTGGGG-ATTTCTCATGCAGCAGCAGC----- | 993 | | | | |
| Qy | 100 | AspArgValIleThrThrSerThrArgThrTrpAlaAlaLeuProThrTyrAsnAsnHisLeu | 119 | | | | |
| Db | 994 | -----GCCACCCACAGCAGGATGGGC----- | 1016 | | | | |
| Qy | 120 | TyrLysGlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyr | 139 | | | | |
| Db | 1017 | -----CAGTTTCCCAAGGCCAAGAGGGCTCAGTCAGGGAAGCCCTTTCATTGCCACC | 1070 | | | | |
| Qy | 140 | SerThrProTrpGlyTyrPheAspPhe----- | 148 | | | | |
| Db | 1071 | TCAGGACCC---GGCCACCTGTGCGCATGCCCCCAGCAGACCCAGCATGGCCCTTCC | 1127 | | | | |
| Qy | 149 | -----AsnArgPheHisCysHis----- | 154 | | | | |
| Db | 1128 | CTGGGCCACCCAGTCGACGACGAGTCTCCACATCACCCTGCTCTCCAGGAGATCC | 1187 | | | | |
| Qy | 155 | -----PheSerProArgAspTrpGlnArgLeuIleAsnAsnAsnTrp | 168 | | | | |
| Db | 1188 | GTTGCCACAGTCACAGATTCTCCCCAACCCCTCTCAACAA-----GGG | 1232 | | | | |
| Qy | 169 | GlyPheArgProLysArgLeuAsnAsnPheLysLeuPheAsnIleGlnValLysGluValThr | 188 | | | | |
| Db | 1233 | GCTGTCAGGCCACAAACCCCTTAATCTTGTTCGGGAACACAGACGTCCCTCCCTACT | 1292 | | | | |
| Qy | 189 | ThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGlnValPheSer | 208 | | | | |
| Db | 1293 | GTAACACAC-----TCA | 1304 | | | | |
| Qy | 209 | AspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCysLeu----- | 226 | | | | |
| Db | 1305 | GGGCAGTATTCTCGATATCCTTAC-----AGTAACCTAAATCAGGAGTATTAGTAACAGT | 1358 | | | | |
| Qy | 227 | -----ProPheProAlaAspValPhe | 234 | | | | |
| Db | 1359 | ACAGGGATGAATCAAAATTTAGGCCTTTACAAACAGTACTCCATGAATCAGTCGGTA--- | 1415 | | | | |
| Qy | 235 | MetIleProGlnTyrGlyTyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSer | 254 | | | | |
| Db | 1416 | -----CCAAGATAC-----CCCAATGCCGTGGGA----- | 1439 | | | | |
| Qy | 255 | SerPheTyrCysLeuGluTyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThr | 274 | | | | |
| Db | 1440 | -----TTCCCGTCA-----AACAGTGGTCAGGACCTCGTG | 1469 | | | | |
| Qy | 275 | PheSerTyrThrPheGluGluValProPheHisSerSerTyrAlaHisSerGlnSerLeu | 294 | | | | |
| Db | 1470 | -----CACCAGCAGCCTTATCCACTCCAGTGGCTCATCTGAACCAAAATGAAC | 1514 | | | | |
| Qy | 295 | AspArgLeuMetAsnProLeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsn | 314 | | | | |
| Db | 1515 | ACACAAATATGACACCT-----TCACAGCCTCAGGA | 1547 | | | | |
| Qy | 315 | GlnSerGlySerAlaGlnAsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMet | 334 | | | | |
| Db | 1548 | ACGTACGCCCTTCCACCTCCCATGTCCACCATGAAGCAATGATGATCCACGAGGC--- | 1604 | | | | |

| | | | |
|-----------------|---|--|------|
| Qy | 335 | SerValGlnProIysAsnTrpLeuProGlyProCytTArGInGlnArgValSerLys | 354 |
| Dd | 1605 | ---ACGCCGCCTCCGAGGTTCAGCGCCGAAG-TGCTGG--- | |
| Qy | 355 | ThrLysThrAspAsnAenSerAenPheThrTrpThrClyAlaSerLysTyAenLeu | 374 |
| Dd | 1640 | -----GATGCCCATGGAGTTGGCAG---TTATCCAATAAT--- | 1672 |
| Qy | 375 | AsnGlyArgGluSerIleAenProGlyThrAlaMetAlaSerHisLysAspAspGlu | 394 |
| Dd | 1673 | ---GCCCCACCCTCACCATCTCACGAGCCCCCTGGTGCCATCGGACAGAGAA | 1729 |
| Qy | 395 | AspLysPhePheProMetSerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSer | 414 |
| Dd | 1730 | -----TATGGGCCCCAGAAAACATGCAGCAGCC | 1756 |
| Qy | 415 | AsnThrAlaLeuAspAsnValMetIleThrAsp- | 430 |
| Dd | 1757 | TGCTTCGTTATTGGGCATGTCTCGGCACCCAGGGAGCTGACCGGACACATGAGACCAA | 1816 |
| Qy | 431 | -----AlaThrAsnProValAlaThrGluArgPhe | 440 |
| Dd | 1817 | TGCTTCCCTGGTTCGGCTTCGGCATCCACAAGCAATCCAGGAACGACTGATACCTGG | 1876 |
| Qy | 441 | GlyThrValAlaValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHis | 460 |
| Dd | 1877 | CCAACAGCAGCCC-----TGGTCAGCAGCCATCTTTCAGCAGTTGCCAACCTGTCC | 1927 |
| Qy | 461 | AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro | 480 |
| Dd | 1928 | TCCACTACAGCCCCACCC-----AGGCCTGCACCA | 1957 |
| Qy | 481 | IleTrpAlaLysIle-ProHisThrAsp-----GlyHisPheHisProSerProLeuMe | 498 |
| Dd | 1958 | GTCTTACCTCCACATCCCATCCCATCACCAGCCTTGGGCACAACTCCACCGTCACCC | 2012 |
| Qy | 498 | tGlyGlyPheGlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValPr | 518 |
| Dd | 2013 | -----CAGAATACCCACACAGAAAGTG-----CCTGTGCA | 2041 |
| Qy | 518 | oAlaAsnProProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSe | 538 |
| Dd | 2042 | TCAGCATTCCTCCATCCGAGCCCTTTCTAGAGAGACCGATGCGGATATGCCAGGTTCAG | 2101 |
| Qy | 538 | rThrGlyGlnValSerVal-----GluIleGlnTrpGluLe | 550 |
| Dd | 2102 | CGCACAGACGCTCAGCTAGTAGAGCGACGATTACTTGCCTCGATAGAGCAGCAGCC | 2161 |
| Qy | 550 | uGlnLysGluAsnSerLysArgTrpAsnProGluVal | 562 |
| Dd | 2162 | ACAGCAGAAGAAGAAAAAGAAAAACCAACCAAT | 2198 |
| RESULT 5 | | | |
| LOCUS | BG207226 | 754 bp mRNA linear EST 21-APR-2001 | |
| DEFINITION | RST26694 Athereys RAGE Library Homo sapiens cDNA, mRNA sequence. | | |
| ACCESSION | BG207226 | | |
| VERSION | BG207226.1 | GI:13728913 | |
| KEYWORDS | EST. | | |
| SOURCE | Homo sapiens (human) | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| AUTHORS | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | |
| | 1 (bases 1 to 754) | | |
| | Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R., | | |
| | Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J. | | |
| | , Lerner,L., Costanzo,D., McGilgott,K., Booser,S., Mays,R., Smith | | |
| | ,E., Veloso,N., Kika,A., Hess,J., Cothren,K., Lo,K., Offenbacher | | |
| | J., Danzig,J. and Ducar,M. | | |
| TITLE | Creation of genome-wide protein expression libraries using random | | |
| | activation of gene expression | | |
| JOURNAL | Nat. Biotechnol. 19 (5), 440-445 (2001) | | |
| MEDLINE | 21227151 | | |

PUBMED 11329013
 CONTACT: Scott J. Cain
 Athersys, Inc.
 3201 Carnegie Ave, Cleveland, OH 44115, USA
 Tel: 216 431 9900
 Fax: 216 361 9596
 Email: scain@atersys.com
 High quality sequence stop: 326.
 Location/Qualifiers
 1..754
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /cell_line="HT1080"
 /clone_lib="Athersys RAGE Library"
 /notes="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."
 223 a 224 c 191 g 115 t 1 others
 BASE COUNT 223 a 224 c 191 g 115 t 1 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.45 Length: 754
 Score: 125.00 Matches: 50
 Percent Similarity: 33.06% Conservative: 31
 Best Local Similarity: 20.41% Mismatches: 68
 Query Match: 3.84% Indels: 96
 DB: 10 Gaps: 11
 US-09-807-802a-15 (1-599) x BG207226 (1-754)
 QY 307 TyTyrLeuAsnArgThrGlnAenGlnSerGlySerAlaGlnAsnLysAspLeuPhe 326
 Db 206 TTTTATCTCAACCAACGGAA-----GCAGGAGGTAAATAATCTAAGGAGCACTGTC 256
 QY 327 SerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyProCys 346
 Db 257 CAAGAAGAGTACAGAAACAGCGCTGGCGTGAATGAGGAACCTGGATG----- 304
 QY 347 TyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsnPheThrTrp 366
 Db 305 -----ACTCGACAGGACCCAGAGCTCTAGATGTTAGATGAGCACTACACTCA 358
 QY 367 ThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThrAla 386
 Db 359 CAAGGA-----AATCTGATTTTCCCTGGGGTTCAC 388
 QY 387 MetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIlePhe 406
 Db 389 TTGGCTCTGATAGCCAGCCAGCATTCCTGGATGGCTTGGCCCTGCTCAGCTAGTG 448
 QY 407 GlyLysGlnSerAlaGlyAlaSerAsnThrAlaLeuAsnValMetIleThrAspGlu 426
 Db 449 GGAGCCGAGACTCTG----- 463
 QY 427 GluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaValAsn 446
 Db 463 ----- 463
 QY 447 PheGlnSerSerThrAspProAlaThrGlyAspValHis---AlaMetGlyAlaLeu 465
 Db 464 -----ACAACACTGCAATGGGGACATTCCTAGTAATGATGGCAAC 508
 QY 466 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 485
 Db 509 AAGGACAGACTG-----GAGCATATCTATCCGGGCC-----GGCCACTG 550
 QY 486 Pro-----HisThrAspGlyHisPheHis----- 493
 Db 551 CCGAAAAACCAACGCTACAAGACACTGCCACCACCGCTGTCAAGTAGGCTCTACTACAA 610

QY 494 -----ProSerProLeuMetGlyGlyPhe 501
 Db 611 CGGAAGCGCGCCGAGGCAAAAAACACCGATGCGCGCCCTAGCGCA----- 664
 QY 502 GlyLeuLysAsnProProGlnIleLeuIleLysAsnThr-----ProVal 517
 Db 665 TAAATATCCGACCCACCCAGATCAAGACAAACCGCACCGGACGAGCAACCCACG 724
 QY 518 ProAlaAsnProPro 522
 Db 725 CCGTGCACCTCCG 739
 RESULT 6
 BHL52154 925 bp DNA linear GSS 27-AUG-2001
 LOCUS ENTPEV12TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
 DEFINITION genomic, genomic survey sequence.
 BHL52154
 ACCESSION BHL52154.1 GI:15315702
 VERSION
 KEYWORDS Entamoeba histolytica
 SOURCE Entamoeba histolytica
 ORGANISM Eukaryota; Entamoebidae; Entamoeba.
 REFERENCE 1 (bases 1 to 925)
 AUTHORS Loftus, B., Wang, Z., Van Aken, S. and Frazer, C.
 TITLE Determination of clone end sequences from Entamoeba histolytica
 JOURNAL HMI:IMSS sheared DNA library (2001)
 COMMENT Unpublished
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjloftus@tigr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Seq primer: M13-Reverse
 Class: shotgun
 High quality sequence start: 12
 High quality sequence stop: 720.
 Location/Qualifiers
 1..925
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: pHO51; Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark, a
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Borell, Oxford University Press, 1999)."
 BASE COUNT 361 a 239 c 96 g 229 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.25 Length: 925
 Score: 121.50 Matches: 86
 Percent Similarity: 32.55% Conservative: 53
 Best Local Similarity: 20.14% Mismatches: 134
 Query Match: 3.74% Indels: 154
 DB: 28 Gaps: 21

KGMNLPDQEMQEVDFLSGFMQSHQSGSLVTRNSTATAMTEQKTGTWQPPQPS
 LPLAPLPQPPQGAHTPLAAQGLAPQSQKQPPQVQYQHLLQPLGSGOOTPPQR
 APKWVHSGSHPSQPMAGSLPLQWFGISLSDLSYSLGSLVSNYMDNMSEVLGOK
 PQGPRNTPWPDQSDGVFGMLGELPLPDPAVGSDFEPARYVAGVSAQMOQROAHH
 RRGNPRGNWPMEDAKRHTWPLPEPFTGDSLSHSGMGAQSDSASSDSTSSANGDSL
 PSMFGSDPLAAVKORHSHSSGEOSTSLPSPLLTWVEDVNDQNKTKWPKAPMOH
 PSPMSTLPSNAPLYAVASPGSQWMDTVQMLQSPWAAANDCNTSTSTYYQTTPQPP
 PPAHKAAPKGFAPFGKAERRPAYLPQ"

polya_signal

5794..5799

/note="putative"

polya_site

5809

/note="putative"

BASE COUNT 1531 a 1471 c 1382 g 1445 t

ORIGIN

Alignment Scores:

| Pred. No.: | 78.6 | Length: | 5809 |
|------------------------|--------|---------------|------|
| Score: | 114.00 | Matches: | 117 |
| Percent Similarity: | 33.57% | Conservative: | 69 |
| Best Local Similarity: | 21.12% | Mismatches: | 201 |
| Query Match: | 3.51% | Indels: | 171 |
| DB: | 11 | Gaps: | 25 |

US-09-807-802A-15 (1-599) x AK076994 (1-5809)

Qy 31 LysLysArgLeuAenPhe-----GlyGlnThrGlyAspSerGlu-----SerVal 45
 Db 2242 GAGAGACGTGAGAACTTCCTGCGTGGGATGATGGCAAGGATGAGAGGCGCACTTA 2301
 Qy 46 ProAspProGlnProLeuGlyGluProAlaThrProAlaValGlyProThrThr 65
 Db 2302 CCCTGATCAGGAATGAGGAA-----GTG 2328
 Qy 66 MetAlaSerGlyGlyAlaProMetAlaAspAsnGluGlyAlaAspGlyValGly 85
 Db 2329 ATAGATTTCTCTCAGGCTTTAAACATGGCCAGTCCCATCAGGGTCCCACTGTGTACA 2388
 Qy 86 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArg----- 101
 Db 2389 AGCGTGAACCTTACTGCG-CACTGCCATGTGTGACTGAACAGAGACAGGAACCATGCAACC 2447
 Qy 102 -----ValIleThrSerThrArgThrTrpAlaLeuProThrTrpAsnAsnHis 118
 Db 2448 ACAGCAGCATCTACTGCCATTGCTCTCCACTACGGCCACCCCGCTGGAGCACAC 2507
 Qy 119 LeuTyrlsGlnIleSer-----SerAlaSerThrGlyAlaSerAsnAsp 133
 Db 2508 GCCTCTGG-CAGCCAGCAAGGTCTGGCTCCTCAGCAGCAGTCTCCAAAGCAGCAGCAGC 2566
 Qy 134 AsnHisTyrlsPheGlyThrProTrpGlyTyrlsPheAspPheAsnArgPheHis--- 152
 Db 2567 CCCAAGTCCAGTACTACAGCACCTCTGTCACCCATTTGATCAGCAGACATCCACCTC 2626
 Qy 153 -----CysHisPheSerProArg 158
 Db 2627 AGCCTCGGGCGCTCGGAATGGGTATCGCTCATCTCAGCACCCATCAGCCCA--- 2683
 Qy 159 AspTrpGlnArg-----LeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsn 176
 Db 2684 ---TGGCAGCGGGTCTGTCTCTCTCTT-GGTCACTGG----- 2715
 Qy 177 PheLysLeuPheAsnIleGlnValLysGluValThrAsnAspGlyValThrThrIle 196
 Db 2716 -----CCTGGCATATCTGATCTC 2733
 Qy 197 AlaAsnAsnLeuThrSerThrValGlnValPheSerAspSerGluTyrlsGlnLeuProTyrl 216
 Db 2734 AGCTCTGACTGTACAGCTTGGGTCTAGTGTAGCAGCTATATGACATATGATGTACAGAA 2793
 Qy 217 ValLeuGlySerAlaHisGlnGlyCysLeuProProPheProAlaAspValPheMetIle 236
 Db 2794 GTTTTGGGACAGAGCCCGGAGGA----- 2817

Qy 237 ProGlnTyrlsGlyTyrlsLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSerPhe 256
 Db 2818 CCTAGAAATAC-----ACCTGGCCAAACCGTCACCAAGTGTATGGA----- 2859
 Qy 257 TyrCysLeuGluTyrlsPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSer 276
 Db 2860 -----GTCTTTTGA 2868
 Qy 277 TyrThrPheGluGluValProPheHisSerSerTyrlsAlaHisSerGlnSerLeuAspArg 296
 Db 2869 ATGTGGGAGAGATTTTGGCTTTTGATCTCTGCATGGGC----- 2907
 Qy 297 LeuMetAsnProLeuIleAspGlnTyrlsLeuAsnArgThrGlnAsnGlnSer 316
 Db 2908 ---TCAGACCCCAAGTTCGCACGCTATGTGCAGAGTGCAGCAGCAATGCACAGAGAAG 2964
 Qy 317 GlySerAlaGlnAsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerVal 336
 Db 2965 CGGCAAGCTCAACATGGCCGC-----CGGCCAGGCACACCCCGTGGC----- 3006
 Qy 337 GlnProLysAsnTrpLeuPro---GlyProCysTyrlsArgGlnGlnArgValSerLysThr 355
 Db 3007 -----AACTGGCCGCCCATCGAAGACGCTCATCGACCTGCGCTTTGCCAGAGTTC 3057
 Qy 356 LysThrAspAsnAsnAsnSerAsnPheThrTrpThrGlyAlaSerLysTyrlsAsnLeuAsn 375
 Db 3058 TTCAGAGAGGGGACAGCTTCACACGCGCTGGTGCAGAGCT-----CAG 3102
 Qy 376 GlyArgGluSerIleIleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAsp 395
 Db 3103 GGAGACTCGGCAGTTCAGTGATGAGACCTCTCTCAGCC-----AATGGGGAC 3150
 Qy 396 LysPhePheProMet-----SerGlyValMetIlePheGlyLys 408
 Db 3151 AGTCTGTTCTCCATGTTTTCAGGCGCTGACCTTGTGCTGTGTTAAACAGAGAGGAAA 3210
 Qy 409 GluSerAlaGlyAlaSerAsnThrAlaLeu-----AspAsnValMetIleThrAspGlu 426
 Db 3211 CACAGCAGTGGAGCAGGAGCAGCAGCAGCTGCCCTCACCACCTCTCTGACCACAGTG 3270
 Qy 427 GluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaValAsn 446
 Db 3271 GAGGACGTGAACACAGGATAACAAACCAAAACG-----TGGCCACCCCAAGACCA 3321
 Qy 447 PheGlnSerSerSerThrAspProAlaThr-----GlyAspValHisAla 461
 Db 3322 TGGCAGACCTTCCCTCCCATGCCAGCAGCTGCCACCCAGCTGCCAGCTCTATGCA 3381
 Qy 462 MetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrlsLeuGlnGlyProIle 481
 Db 3382 GTC---GCCAGTCTCTGGCAGCAGTGGAAACGACACCGTGCAGATGCTGCAGTCCCTGTG 3438
 Qy 482 TrpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPhe 501
 Db 3439 TGGGCTGACGCAATGACTGC----- 3459
 Qy 502 GlyLeuLysAsn-ProProGlnIleLeuIleLysAsnThrProValProAlaAsnPr 521
 Db 3460 -----AATACCACTCTCTTACC-TATGTTGACAGCACCCACACAGCCCCCCTCC 3509
 Qy 521 oProAlaGluPheSerAlaThrLys---PheAlaSerPhe 533
 Db 3510 ACCAGCACATAAGGCAGCAGCACCCAGGGCTTCAAGGCTTC 3549

RESULT 10

BC037597

LOCUS

DEFINITION

BC037597

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

BC037597 2637 bp mRNA linear HTC 04-MAR-2003
 Mus musculus, Similar to plakophilin 4, clone IMAGE:5250198, mRNA.

BC037597.1 GI:22902283

HTC.

Mus musculus (house mouse)

Mus musculus


```

Db      1084 CA-----GAGACCGCGCTCCCAACAGCAGCGCGCGCGGTCTGTCTACCTCC 1134
Qy      444 laValaAsnPhelGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyA 464
Db      1135 CGGAGACCTCCATCCCAACGAGCAGTCCCTCAGTACCAGACACACACAGGCGTGGG 1194
Qy      464 la-----
Db      1195 TCCCACTGACCTGACTGATGCCAGACTCGAGTAGTCTTCCCATCCCAAGGCCAGGTG 1254
Qy      465 -----LeuProGlyMetValTrpGlnAspArg-----
Db      1255 GGGTCATGTCCTCCGAAACGCTTGGGATGACCGCGGTACACAGCAGCCTGGGACCTTCA 1314
Qy      474 -----AspValTrpLeuGlnGlyProIleTrp---AlaLysIleProHis-----T 488
Db      1315 CTGCAAGGAGCTGTTTCATGACATGGACCAATTTGGACAGCAGCAGTATGACATTTATGAA 1374
Qy      488 hrAspGlyHisPhe-HisProSerProLeuMetGly-----
Db      1375 AGGATGTTTCCACCTCGGCGCAGACAGCTTGACAGGCTTACGGAGTTCCTATGCCAGTCAG 1434
Qy      500 -----GlyPheGlyLeuLysAsnPro---ProProGlnIleLeuIleLysAsn 514
Db      1435 CATAGTCAGCTTGGGCAAGAGCTCCGTTTCAGCTGTGTCTCTGACTTGACATC----- 1488
Qy      515 ThrProVal-----
Db      1489 ACTCCCTATCATGAAGGAGGAGCTACTACAGTCCAGTGTACCGCAGCCCAACACAGGG 1548
Qy      522 ProAlaGlnPheSerAlaThrLysPheAlaSerPheIleThrGlnTrpSerThrGlyGln 541
Db      1549 ACCGTGGAGCTCCAGGGGTACAGACAGCA-----TTGTATCCGACAGCTCA 1596
Qy      542 ValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGlu 561
Db      1597 GTAGGTATT-----GGAAATCTACAAAGACATCCAGCAGCA----- 1635
Qy      562 ValGlnTrpThrSerAsnTrpAlaLysSerAlaAsnValAspPheThrValAspAsnAsn 581
Db      1636 -----AGCACCTTACATACCAAGAAT-----AACTATGCCCTGAACACAGCA 1680
Qy      582 GlyLeuTrpThrGluPro---ArgProIleGlyThrArg 593
Db      1681 GCTACCTATCGGAGCCCTACAGGCCAGTTCAGTACCGA 1719

```

RESULT 11

```

BC019128          2688 bp  mRNA  linear  HTC 20-SBP-2002
LOCUS             Mus musculus, Similar to plakophilin 4, clone IMAGE:5008935, mRNA.
DEFINITION
ACCESSION         BC019128
VERSION           BC019128.1 GI:17512314
KEYWORDS          HTC.
SOURCE            Mus musculus (house mouse)
ORGANISM          Mus musculus
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
                  (bases 1 to 2688)
REFERENCE
AUTHORS           Strausberg,R.
TITLE             Direct Submission
JOURNAL           Submitted (07-DEC-2001) National Institutes of Health, Mammalian
                  Gene Collection (MGC), Cancer Genomics Office, National Cancer
                  Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
                  USA
REMARK            NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT           Contact: MGC help desk
                  Email: gcapsb-remail.nih.gov
                  Tissue Procurement: Gilbert Smith, Ph.D.
                  cDNA Library Preparation: Life Technologies, Inc.
                  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                  DNA Sequencing by: National Institutes of Health Intramural
                  Sequencing Center (NISC),

```

Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@nhgri.nih.gov
 Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
 Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooke,S.,
 Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
 Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
 Maduro,Q.L., Masello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
 McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
 Tourgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
 Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 39 Row: e Column: 15
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis, similarity but not identity to protein
 This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers
 source

```

1..2688
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:5008935"
/tissue_type="Mammary tumor, MMTV-LTR/INT3 model. 5 month
old mouse. Taken by biopsy."
/clone_lib="NCI CGAP Mam2"
/lab_host="DH10B"
/notes="Vector: pQWV-SPORT6"
BASE COUNT      693 a 736 c 726 g 533 t
ORIGIN

```

Alignment Scores:

```

Pred. No.:      29.8      Length:      2688
Score:          113.50    Matches:     148
Percent Similarity: 34.96% Conservative: 81
Best Local Similarity: 22.60% Mismatches: 228
Query Match:      3.49% Indels:      199
DB:              11      Gaps:       37

```

US-09-807-802A-15 (1-599) x BC019128 (1-2688)

```

Qy      21 GlyIleGlyLysThrGlyGlnGln-----ProAlaLysValArgLeuAsnPheGlyGln 38
Db      115 GGGAGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 174
Qy      39 ThrGlyAspSerGluSerValProAspProGlnPro-LeuGlyGluProProAlaThrPr 58
Db      175 TTTGCT-----CCGGTGGCGAGCGGCGGAGCGGCGGAGCGGCGGCGGCGGAT 213
Qy      58 oAlaAlaVal---GlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAs 77
Db      214 GTCCCTGTGTCCTGGAGCGACGACGCGCTCCGCTGAGCTGAGAAAGAGGAATGCGCGCCC 273
Qy      77 nAsnGluGlyAlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTr 97
Db      274 CTGAACAGGGCTCACTGTGTGGAGAGGGGCA-----
Qy      97 pLeuGlyAspArgValIleThrThr-----SerThrArgThrTrpAlaLeuPr 113
Db      305 -----ACCACAGACCCACAGGAGCTGTCTCCACTGG-----CC 339
Qy      113 oThrTrp-----AsnAsnHisLeuTrpLysGlnIleSerSerAlaSerThrGln 129
Db      340 CAGGCATGGAAACCCGAGACACACAGCCACCATCTATTCTAGCTTCGTTGAGAGGAGCA---GG 396
Qy      129 yAlaSerAsnAspAsnHisTrpPheGlyTrpSerThrProTrpGlyTrpPheAspPheAs 149
Db      397 AGCTTCA-----GTTTCACAGCTACCCCGAGAACTGGAGTGG 435
Qy      149 nArgPheHisCysHisPheSerProArgAsp-----TrpGlnArgLeuIleAsnAs 166

```



```
Db 436 AAAGCAGATTGTTGCCAGTCAGCTAGAAAGATGTAGGCTTGGAGCAGAAATCCGCAAGCA 495
Qy 166 nAsnTrpGlyPheArgProLysArgLeuAsnPhelYsLeuPheAsnIleGlnValLys-G 186
Db 496 T-----CGCCAGCACAGCTCAACTGAAAGTCAATTCCTTTGGAGATCAACAG 543
Qy 186 luValThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGlnV 206
Db 544 ATGTG-----CCAAATCTGTGTG-----AGCAAACTTAGAG 576
Qy 206 alPheSerAspSerGluTrpGlnLeuProTyrValLeuGlySer---AlaHisGlnGlyC 225
Db 577 TT---TCTGACACTATTATCCCAACAACTATCATCAGCAGCAGCAGCAGCAGGGA 633
Qy 225 ysLeuProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuA 245
Db 634 CCTTC-----TACTCACCAGAACAGACTTCTCTCCATGAAA 669
Qy 245 snAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerG 265
Db 670 GTGAGGATCGTTGGAACTCAAGAGTTCA-----ACAC 705
Qy 265 lnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheH 285
Db 706 AGATG-----AATCTTACTCTGACAGTGGGTACAGGAGCAGGAGGAGTTTCC 753
Qy 285 isSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnT 305
Db 754 ACAACGCAACCGTGAGCAAGGCA---GACAGCAGACAGCATCCATTCACAGGATCAA 810
Qy 305 yrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuL 325
Db 811 CAAGTAACCATGTGTGAGGACTTCAAGAGCTGAGGACAAACACTGTTTCAGCCATCAG 870
Qy 325 euPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyP 345
Db 871 TAGCAATCGGGCCATCGCGAGAGTAGTTAGTTCCATCTAGAGCAGCAGTCTCTCTT 930
Qy 345 roCysTyrArgGlnGlnArgValSer---LysThrLysThrAspAsnAsnSerAsn 364
Db 931 A-TGTTACCACACAGCGGTCTCTCTCAAGGGGTCACTGAGAATCTCTCTGGTAGT 989
Qy 364 heThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleLeuAsnProG 384
Db 990 GGATTGGCTCTCCATCAGT-----GACGGACTCCCGACC---TCTGAACCCC 1034
Qy 384 lyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValM 404
Db 1035 AGTGCTTACTCTCCAGCAC----- 1054
Qy 404 etIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleT 424
Db 1055 -----GTCCTCTGCACAGCGGGCGGCTCTCCATATCTCA 1088
Qy 424 hrAspGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValA 444
Db 1089 CA-----GAGACCGGCTCTCCCAACAGCGCGTGGCGGGTCTGTCTGTCTCACCTCC 1139
Qy 444 laValAsnPhelGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyA 464
Db 1140 CGGCAGACCTCCAAATCCCAACGGACCACTGCTCTAGTACCAGACCAACCGAGGGTGGG 1199
Qy 464 la----- 464
Db 1200 TCCCCACTGACCTGACTGATGCCAGACTCGAGTAGTCTCCCATCCCAAGGCCAGGTG 1259
Qy 465 -----LeuProGlyMetValTrpGlnAspArg----- 473
Db 1260 GGGTCATCGTCCCGAAACGCTCTGGGATGACCGCGGTACCACAGCAGCCTGGGACCTTCA 1319
Qy 474 -----AspValTyrLeuGlnGlyProIleTrp---AlaLysIleProHis-----T 488
```

```
Db 1320 CTGCAAAAGGAGTCTGTCATGATGACGACCAATTTGGACAGCAGCAGTATGATTTATGAA 1379
Qy 488 hrAspGlyHisPhe-HisProSerProLeuMetGly----- 499
Db 1380 AGGATGGTTCCACTCGGCGCAGCAGCGCTTACAGCGCTTACGAGGTTCTTATGCGAGTCAG 1439
Qy 500 -----GlyPheGlyLeuLysAsnPro---ProProGlnIleLeuIleLysAsn 514
Db 1440 CATAGTCAGCTGGGCAAGAGCTCCGTTACGTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1493
Qy 515 ThrProVal-----ProAlaAsnPro 521
Db 1494 ACTCTCTATCTATGAAGGAGGAGCCTACTACAGTCCAGTACCGCAGCAGCCCAACCCAGGG 1553
Qy 522 ProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGln 541
Db 1554 ACCGTGGAGTCCAGGGGTACAGACAGCA-----TTGATCCGACAGGCTCA 1601
Qy 542 ValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGlu 561
Db 1602 GTAGGTATT-----GGAAATCTACAAAGGACATCCAGCCAGCGA----- 1640
Qy 562 ValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsn 581
Db 1641 -----AGCACCTTATACAAAGAAAT-----AACTATGCCCTGAACACAGCA 1685
Qy 582 GlyLeuTyrThrGluPro---ArgProIleGlyThrArg 593
Db 1686 GTACCTATCGGAGCGCTACAGCGCAGTTTCAGTACCGA 1724
```

RESULT 12
AY103595
LOCUS Zea mays PC0108378 mRNA sequence. 1960 bp linear HTC 16-OCT-2002
DEFINITION Zea mays PC0108378 mRNA sequence.
ACCESSION AY103595
VERSION AY103595.1 GI:21206673
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 1960)
Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S.,
Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
Unpublished (2002)
REFERENCE 2 (bases 1 to 1960)
Coe, E.H.
Direct Submission
Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones,
these are publicly available from ZmDB and may be found by BLAST
searching at MBL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
maize cDNA sequences is either Virginia Walbot, Stanford or Pat
Schnable, Iowa State, then clones may be requested from ZmDB:
www.zmdb.iastate.edu.
Location/Qualifiers
1. 1960
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:636289"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Consensus
Library"
/notes="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize

FEATURES
source

BASE COUNT 497 a 461 c 489 g 513 t
ORIGIN

Alignment Scores:

Pred. No.: 21.3 Length: 1960
Score: 113.00 Matches: 117
Percent Similarity: 31.02% Conservatives: 75
Best Local Similarity: 18.90% Mismatches: 235
Query Match: 3.48% Indels: 192
DB: 11 Gaps: 24

US-09-807-802A-15 (1-599) x AV103595 (1-1960)

```

QY 17 AspSerSerGlyIleGlyThrGlyGlnGlnProAlaIleValysArgLeuAsnPhe 36
   |||||
DB 34 GACAGTCTGAGCGCTGCGGGGAGTCTAACAGCACCCCTCTCGCTCGCACTC----- 87
   |||||

QY 37 GlyGlnThrGlyAspSerGluSerValProAspProGlnProLeuGlyGluProAla 56
   |||||
DB 88 ----CGTTCGTCTGACATCTCTCCCGTCTCTTCTCTCTGAGGCTCGGACCCGACAA 144
   |||||

QY 57 ThrProAlaAlaValGlyProThrMetAlaSerGlyGlyGlyAlaProMetAlaAsp 76
   |||||
DB 145 TGCGCCCAACCGCGGTGCGTGCAGCAGAGAGCTCGCAAGCTTCGCGCGAGGTGCGCA 204
   |||||

QY 77 AsnAsnGluGlyAlaAspGlyValGlyAsnAlaSerGlyAsnTrpHisCysAspSerThr 96
   |||||
DB 205 AACTCAGCCAGATCAGCAGAGAGAGAGCGGGTTCATCAGCCCTCGTGCAGCTACC 264
   |||||

QY 97 TrpLeuGlyAspArgValIleThrThrSerThrArgThr----- 109
   |||||

DB 265 TCAGTGGGAGCGGAGCAGATCGAGTGGAGCAGATCCAGACCCCGAGGTAG 324
   |||||

QY 110 TrpAlaLeuProThrTyrAsnAsnHisLeuTyrLys----- 121
   |||||

DB 325 TGGTGGCGTACGATACCTCAGCTCGCTCTCGAGATCTCGAGGAGAGAGAGAGTGC 384
   |||||

QY 122 ----GlnIleSerSerAlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGly 138
   |||||
DB 385 TGGACAAAGCTGCTGTGCTCAAGCTCAACGAG-----GGC 420
   |||||

QY 139 TyrSerThrProTrpGly-TyrPheAspPheAsnArgPheHisCysHisPheSerProAr 158
   |||||
DB 421 TCGGGACGACCATGGGCTGCAC-----CGGACCCAGTCTGTCATTTGAAGTCGCA 471
   |||||

QY 158 g-----AspTrpGlnArgLeu-IleAsnAsnAsnTrpG 169
   |||||
DB 472 ATGGGTTACATCTCTTACCTTATTTGATTCATAATCGAGTCCCTGAAACAGAGATG 531
   |||||

QY 169 lypheargProLysArgLeuAsnPhelysLeuPheAsnIleGlnValLysGluValThrT 189
   |||||
DB 532 GATGTAATGTCCTTACTTCTGATGAACTCTTCTCAAT-----A 570
   |||||

QY 189 hrAsnAspGlyValThrThrIleAla-----AsnAsnLeuThrSerThrValGlnV 206
   |||||
DB 571 CCCATGATGACACACAGAGATCGTTGAGAAGTATTCCTCAACTCCAACTTGAATTCATA 630
   |||||

QY 206 alPheSerAspSerGluTyr-----GlnLeuProTyrV 217
   |||||
DB 631 CTTTTCATCAGACCGCAGTATCTCGCATTTTACCAGGAGCTTCTTGGCCACTTCCCAAGCA 690
   |||||

QY 217 alLeuGlySerAlaHisGlnGlyCysLeuProPheProAlaAspValPheMetIleP 237
   |||||
DB 691 AAGGGAATCTGGGAAGATGGCTGGTATCTCAGGCCATGGTGATGTTCTCC----- 746
   |||||

QY 237 roGlnTyrGlyTyrLeuThrLeuAsnAsnGlySerGlnAlaValAlGlyArgSerSerPheT 257
   |||||
DB 747 -----TCCTTGAATAACAGTGGAAACTC----- 770
   |||||

QY 257 yrCysLeuGluTyrPheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerT 277
   |||||
DB 771 -----GACATCTTATTTGGTCTCAGGGCAAGGAGTATGTCTTCTC----- 806
   |||||

```

```

QY 277 yrThrPheGluGluValProPheHisSerSerTyrAlaHisSerGlnSerLeuAspArgL 297
   |||||
DB 807 -----GTTCTTAACACACAACTTGGGTGCTA 834
   |||||

QY 297 euMetAsnProLeuIleAspGlnTyrLeuTyrTyrLeuAsnArgThrGlnAsnGlnSerG 317
   |||||
DB 835 TAGTCGACATCAAGATC-----CTGAACCATCTGATCAATAACAGAAATGAA----- 881
   |||||

QY 317 lySerAlaGlnAsnLysAspLeuPheSerArgGlySerProAlaGlyMetSerValG 337
   |||||
DB 881 ----- 881
   |||||

QY 337 lnProLysAsnTrpLeuProGlyProCysTyrArgGlnArgValSerLysThrLysT 357
   |||||
DB 882 -----TACTGCATGGAGGTACTCCAAAACATTTGG 912
   |||||

QY 357 hrAspAsnAsnAsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyA 377
   |||||
DB 913 CTGATGTTAAA-----GGCGGTACTCTCATCTCTTACGAAGNA 951
   |||||

QY 377 rgGluSerIleIleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysP 397
   |||||
DB 952 GAGTTTCAGCTTTTGGAGATTGCCCAAGTACTGATGAGCATGTGAATGAG-----T 1002
   |||||

QY 397 hePheProMetSerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrA 417
   |||||
DB 1003 TTAATCATCGAGAGTTTAAGATATTTCAACACTTACCACTTGGGTGAACCTTAAAG 1062
   |||||

QY 417 laLeuAspAsnVal-----MetIleThrAspGluGluG 428
   |||||
DB 1063 CTGTCACAGAGACTAGTAGAGCTGAGGCTGAGGCACTTAAGATGGAATATTATCCAAACCCCAAG 1122
   |||||

QY 428 luIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaValAsnPheG 448
   |||||
DB 1123 AAGTTGATGGTGTGAAGTCTTCAACTTGAACACTGACAGCTGGTGCAGCTATTCTGTTTC- 1181
   |||||

QY 448 lnSerSerThrAspProAlaThrGly---AspValHisAlaMetGlyAlaLeuProG 467
   |||||
DB 1182 -----TTTGACAAAGCGATTGGAATTAATGTTCCCGCTCAAGATTTCTCCCGG 1230
   |||||

QY 467 lyMetValTrpGlnAsp-----ArgAspValTyrLeuGlnGlyProIleT 482
   |||||
DB 1231 TGAAGGCTACATCTGATTTATTTGCTGTCAGTCTGTATCTTTAC----- 1274
   |||||

QY 482 tpAlaLysIleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheG 502
   |||||
DB 1275 -----ACCTTGGTTGATGGCTTTG 1293
   |||||

QY 502 lyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsnProp 522
   |||||
DB 1294 TCATCCGCAATCA-----TCCAGAGCAATCCAGCTAACCTTT 1332
   |||||

QY 522 roAlaGluPheSerAla-----ThrLysPheAlaSerPheIleThrGlnTyrSerThrG 540
   |||||
DB 1333 CGATGAGCTGGACCTGAGTTCAAGAAGTTGCCAATTTCTGCTCGTTCAAGTCCA 1392
   |||||

QY 540 lyGlnValSerValGluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnP 560
   |||||
DB 1393 TCCCGAGCATCTGCGAGCTTGAC-----AGCTTGAAGTTTCTG 1431
   |||||

QY 560 roGluValGlnTyrThrSerAsnTyrAlaLysSerAlaAsnValAspPheThr 577
   |||||
DB 1432 GTGATGCTCTGTTGTTCTGGAATTCAGCTCAAGGGCAAGGTGACAAATCACC 1484
   |||||

```

RESULT 13

AK048546

LOCUS

DEFINITION

AK048546 4782 bp mRNA linear. HTC 05-DEC-2002
Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched
library, clone: C130072F01 product: COLLAGEN ALPHA 1(III) CHAIN

PRECUSOR, full insert sequence.

AK048546

ACCESSION

VERSION

AK048546.1 GI:26339397

Percent Similarity: 34.97% Conservative: 49
Best Local Similarity: 23.54% Mismatches: 175
Query Match: 3.48% Indels: 104
DB: 11 Gaps: 27

US-09-807A-15 (1-599) x AK048546 (1-4782)

QY 4 GlyLeuValArgProVal---GluGlnSerProGlnGluProAspSerSerSerSergly--- 21
DB 3524 GTGCCAGAGGCACGATTTGGACCACATCGCATCTCCTGGAAAAGATGGAACAAGTGGCCAT 3583
QY 22 IleGlyThrGlyGlnGlnProAlaLysValSerGluLeuAsnPheGlyGlnThrGlyAsp 41
DB 3584 CAGGTCTCTATTGGA-----CCACCGGTCTTAGAGAAACAGAGGTGAAGAGGATCT 3637
QY 42 SerGluSerValProAspProGlnProLeuGlyGluProProAlaThProAlaLaVal 61
DB 3638 GAGGGCTCCGAGCCACCTCGACAGCCAGCA--CCCCCTGGACCCCTTGTCGCCCT 3694
QY 62 GlyProThrThrMetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAla 81
DB 3695 GGTCCC-----TGCTGGTGGTGGTCTGCTGCCATGTCT----- 3730
QY 82 AspGlyVal---GlyAsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAsp 100
DB 3731 ---GGAGTTGGAGGTGAAGAAGCTGGTGCCTTT-----TCACCCCTATTATGGAGAC 3778
QY 101 Arg-----ValIleThrThrSerThrArgThrTrpAlaLeuProThrThrAsn 116
DB 3779 GATCCAATGGATTTCAGATCAACACTGAAGAGATTATGCTTCACTCAAGCTCTGTTAAT 3838
QY 117 AsnHisLeuTyrlsGlnIleSer-----SerAlaSerThrGlyAlaSerAsnAsp 133
DB 3839 GGCAAATAGAGAGCTTTATAGCCCTGATGTTCTCGAAAAAACCCCTGCTCGAGTGC 3898
QY 134 AsnHieTyrlsPheGlyTyrSerThrProTrpGlyTyrPheAspPheAsnArgPheHisCys 153
DB 3899 AGAGACCTA-----AAATTC-----TGC 3916
QY 154 HisPheSerProArgAspTrpGlnArgLeuIleAsnAsnAsnTrpGlyPheArgProlys 173
DB 3917 CACCCGAAGTCAAGAGTGGAGATACTGGTGTGATCTTACCAAGGCTGCAAGTGGAT 3976
QY 174 ArgLeuAsnPhelysLeuPhe---AsnIleGlnValLysGluValThrAsnAspGly 192
DB 3977 GCTATA-----AAAGTATTCTTAACATAGGAACCTGGGAAACATCATATAATGCCAGC 4030
QY 193 ValThrThrIleAlaAsnLeuThrSerThrValGlnValPheSerAspSerGluTyr 212
DB 4031 CCCATGACTCTCCACGTAAG-----CACTGGTGACAGATTCTGGTGCA 4075
QY 213 GlnLeuProTyrVal---LeuGlySerAlaHisGlnGlyCysLeu----- 226
DB 4076 GAGAGAATAATGTTGGTTTGGAGAACTATCAATGGTGGTTTTTCAGTTCAGCTATGGC 4135
QY 227 ---ProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 245
DB 4136 CCTCCTGATCTCTCGAAGATGTCGTTGATGTG---CAGCTGGCATCTCTCAGACTTCT 4192
QY 246 AsnGlySerGlnAlaValGlyArgSerSerPhePheTyrCysLeuGluTyrPheProSerGln 265
DB 4193 TCC---AGCCGGGCTCCGAGACATATACATCACCTGCAAAAACAGTATGTCCTATATG 4249
QY 266 MetLeuArgThrGlyAsn----- 271
DB 4250 GATCAGGCCAGTGGCAATTAAGAAAGTCTCTGAAGCTGATGGATCAAAATGAAGCGAA 4309
QY 272 -----AanPheThrPheSerTyrThrPheGluGlu----- 281
DB 4310 TTCAAGGCTGAAGGAACACGAANAATTCATTCACAGTTCTTAGAGAGTGGCTGTACTAAA 4369
QY 282 -----ValProPheHisSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsn 299

| | | | | | | |
|----|------|-------------------------|--|-------------------------------------|--------------------------------|-----|
| Qy | 251 | ValGlyArgSerSerPhe | ThrCysLeu | GluThrPheProSerGlnMetLeu | ArgThrGly | 270 |
| Db | 758 | GTGACTCCTTT | CATACAAGACAACA | GAGAAGAAAGAAAGACG | CACCTGCTACCAAG | 817 |
| Qy | 271 | AsnAsn | PheThrPhe | SerTyrThr--Phe | GluGluValProPheHisSerSerTyrAla | 289 |
| Db | 818 | GGTGGGTTCACAT | TGGTAAGTGGGTCTCTCTTTTGGCATCTTCCTCAATGTTGTT | TTGTT | 877 | |
| Qy | 290 | HisSerGlnSerLeu | AspArgLeuMetAsnProLeu | IleAspGlnTyrLeuTyrTyrLeu | 309 | |
| Db | 878 | TTGGGAAGGACAGAGAGAGA | ACCAACCAACCTGTATCTTCTCTGGGTGTTGGG | 937 | | |
| Qy | 310 | AsnArgThrGln | AsnGln-- | -----SerGlySerAla | 319 | |
| Db | 938 | AAGAAACCCGACAG | GTGAGGACCAAGTGTCAAGCAGTATTTCTTTGGGAATTCAGAG | 997 | | |
| Qy | 320 | GlnAsn | LeuAsp-- | -----LeuLeuPheSerArgGlySerProAlaGly | 333 | |
| Db | 998 | CAAACCAAGATGAGAGCT | CAAGCCGACATTCAGCTGTGGGCAAAACCATCTGGG | 1057 | | |
| Qy | 334 | MetSerValGlnPro | LysAsnTrpLeuProGlyProCysTyrArgGlnGlnArgValSer | 353 | | |
| Db | 1058 | AAGGAG-- | -----TCTGAGCAGCTAGCAAG | 1081 | | |
| Qy | 354 | LysThrLysThrAspAsn | AsnAsnSerAsn | PheThrTrp--ThrGlyAlaSerLysTyr | 372 | |
| Db | 1082 | GCTACTTTTGCATTTGGAAATCA | AAACCAATACAACTGATCAAGGTGCAGCAAGCCA | 1141 | | |
| Qy | 373 | AsnLeuAsnGlyArgGlu | SerIleAsnProGlyThrAlaMetAlaSerHisLysAsp | 392 | | |
| Db | 1142 | GTTTTTAGCTTCTGA | CAGCAGTTCCTCTAGTTCAAGTGCACGCACT-- | 1192 | | |
| Qy | 393 | AspGluAspLysPhePhePro | MetSerGlyValMetIlePheGlyLysGluSerAlaGly | 412 | | |
| Db | 1193 | ----- | TCATAGTGGTGGCATATTTGCT-- | 1216 | | |
| Qy | 413 | AlaSerAsnThrAlaLeu | AspAsnValMetIleThrAspGluGluIleLysAlaThr | 432 | | |
| Db | 1217 | --AGTTCCACCTCTCTC | CAACCCACCTGTGGCAGCCTTGTGTTCAGCAGGCGACG | 1273 | | |
| Qy | 433 | AsnProValAlaThrGlu | ArgPheGlyThrValAlaValAsnPheGlnSerSerSer-- | 451 | | |
| Db | 1274 | AATCTGTGCAGCAGCTCGCT | TCGGTAAACGCTGCTGATCCAGTACATCTCAGTCTTTG | 1333 | | |
| Qy | 452 | ----- | ThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeuPro | 466 | | |
| Db | 1334 | CTATTTCTCAAGAGAGCGAGCC | AGCCACACGTCACAGCTGCCCGCAGCTGCTCTCC | 1393 | | |
| Qy | 467 | GlyMetValTrpGln | AspArgAsp-- | -----ValTyrLeuGln | 478 | |
| Db | 1394 | TT-TGTATTGG-- | GACAGAGCGCAGCATATAGTGTCTCTGGGTTCACCTTTTGG | 1449 | | |
| Qy | 479 | GlyProIleTrpAla | LysIle--ProHisThrAsp-- | -----GlyHisPheHi | 493 | |
| Db | 1450 | GGCTACAACCAACGTC | CAAGCTCTTCAGGATCTCTCTTTGATTTGGCACTGGACATTCAGC | 1509 | | |
| Qy | 493 | sProSer-- | -----ProLeuMetGly-- | -----G1 | 500 | |
| Db | 1510 | ACCATCTGCCAGTC | CCAGCATTTGGTGTCTAACCAAGTCTCAACATTTGGCAAAAGTCAAGG | 1569 | | |
| Qy | 500 | yPheGlyLeuLysAsnPro | ProProGlnIleLeuIleLysAsnThr-- | ----- | 515 | |
| Db | 1570 | TGCCAGCCAGCCTAAC | CGCCAAAGCTTTGGCTCTATCTCGTCTTCAACAGCATATTATTTTC | 1629 | | |
| Qy | 516 | ----- | ProValPro-- | -----AlaAsnPr | 521 | |
| Db | 1630 | TGCTGGTTCTCAGCCTGT | ACCACCTATATTTTGGGCAGTGTCAAGCAGCAGCCCAACC | 1689 | | |
| Qy | 521 | oPro-- | -----AlaGluPheSerAlaThrLysPheAlaSe | 532 | | |
| Db | 1690 | TCCTGTGTGTGGACAGCAGC | CTAGTCTGCAATTTGGTCTCGGACAGCATGATGCAG | 1749 | | |

Search completed: February 15, 2004, 01:28:59
Job time : 3387.15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 14, 2004, 15:19:20 ; Search time 5428 Seconds
(without alignments)
4024.639 Million cell updates/sec

Title: US-09-807-802A-17
Perfect score: 2906
Sequence: 1 MASGGAPNADNNEGADGVG.....NNGLYTEPRPICTRYLTRPL 534

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2886711 seqs, 2045481336 residues
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DRV=xlh
-Q/cgn2_1/USPTO_spool/US09807802/runat_11022004_175607_15929/app_query.fasta_1.2389
-DB=GenEmbl -QFWT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09807802 @CGN 1.1 12216 @runat_11022004_175607_15929 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.ets.*
12: gb.ey.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rpd.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.ey.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|--------|---------------|--------|-------|------------------------|
| 1 | 2306 | 100.0 | 4718 | 14 | AF063497 Adeno-888 |
| 2 | 2884 | 99.2 | 4683 | 14 | AF028704 Adeno-888 |
| 3 | 2530 | 87.1 | 4722 | 14 | AF028705 Adeno-888 |
| 4 | 2528 | 87.0 | 4726 | 14 | AF028704 Adeno-888 |
| 5 | 2518 | 86.6 | 8179 | 6 | AR264580 Sequence |
| 6 | 2481.5 | 85.4 | 4072 | 6 | AR222046 Sequence |
| 7 | 2481.5 | 85.4 | 4072 | 6 | AX205074 Sequence |
| 8 | 2481.5 | 85.4 | 4679 | 6 | AX282480 Sequence |
| 9 | 2481.5 | 85.4 | 4679 | 14 | AF043303 Adeno-888 |
| 10 | 2481.5 | 85.4 | 7327 | 12 | AF369963 Cloning v |
| 11 | 2481.5 | 85.4 | 7557 | 6 | AR222045 Sequence |
| 12 | 2481.5 | 85.4 | 7557 | 6 | AX205073 Sequence |
| 13 | 2481.5 | 85.4 | 8698 | 6 | AR222044 Sequence |
| 14 | 2481.5 | 85.4 | 8698 | 6 | AX205072 Sequence |
| 15 | 2457.5 | 84.6 | 2208 | 14 | AY243021 Non-human |
| 16 | 2439.5 | 83.9 | 4680 | 6 | AR028767 Sequence |
| 17 | 2439.5 | 83.9 | 4680 | 6 | IG2303 Sequence 1 |
| 18 | 2439 | 83.9 | 2187 | 14 | AY243013 Non-human |
| 19 | 2437.5 | 83.9 | 2196 | 14 | AY243010 Non-human |
| 20 | 2436 | 83.8 | 2187 | 14 | AY243006 Non-human |
| 21 | 2436 | 83.8 | 2187 | 14 | AY243017 Non-human |
| 22 | 2431.5 | 83.7 | 2196 | 14 | AY243011 Non-human |
| 23 | 2430 | 83.6 | 4721 | 14 | AF513851 Adeno-888 |
| 24 | 2429.5 | 83.6 | 2217 | 14 | AY243015 Non-human |
| 25 | 2428.5 | 83.6 | 2217 | 14 | AY243009 Non-human |
| 26 | 2427.5 | 83.5 | 2217 | 14 | AY243007 Non-human |
| 27 | 2425 | 83.4 | 2187 | 14 | AY243016 Non-human |
| 28 | 2423.5 | 83.4 | 2202 | 14 | AY243012 Non-human |
| 29 | 2423.5 | 83.4 | 4675 | 6 | AX135805 Sequence |
| 30 | 2423.5 | 83.4 | 4675 | 6 | AX286292 Sequence |
| 31 | 2423.5 | 83.4 | 4675 | 6 | BD094552 Method of |
| 32 | 2423.5 | 83.4 | 4675 | 14 | AA2CG J01901 Adeno-888 |
| 33 | 2422.5 | 83.4 | 2202 | 14 | AY243004 Non-human |
| 34 | 2422 | 83.3 | 2187 | 14 | AY243019 Non-human |
| 35 | 2421.5 | 83.3 | 2202 | 14 | AY243014 Non-human |
| 36 | 2421.5 | 83.3 | 2217 | 14 | AY243023 Non-human |
| 37 | 2419.5 | 83.3 | 2190 | 14 | AY242998 Non-human |
| 38 | 2418.5 | 83.2 | 2190 | 14 | AY242999 Non-human |
| 39 | 2418 | 83.2 | 2187 | 14 | AY243018 Non-human |
| 40 | 2417 | 83.2 | 2214 | 14 | AY243020 Non-human |
| 41 | 2411.5 | 83.0 | 2190 | 14 | AY243000 Non-human |
| 42 | 2411.5 | 83.0 | 2217 | 14 | AY243022 Non-human |
| 43 | 2410 | 82.9 | 2208 | 14 | AY243008 Non-human |
| 44 | 2410 | 82.9 | 2211 | 14 | AY242997 Non-human |
| 45 | 2390 | 82.2 | 1932 | 14 | AY243024 Non-human |

ALIGNMENTS

```

AF063497      4718 bp      DNA      linear      VRL 27-APR-1999
LOCUS      Adeno-associated virus 1, complete genome.
DEFINITION      AF063497
ACCESSION      AF063497
VERSION      AF063497.1 GI:4689096
KEYWORDS      Adeno-associated virus 1
SOURCE      Adeno-associated virus 1
ORGANISM      Adeno-associated virus 1
REFERENCE      Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
AUTHORS      1 (bases 1 to 4718)
Xiao.W., Chirmule.N., Berta.S.C., McCullough.B., Gao.G. and
Wilson.J.M.
TITLE      Gene therapy vectors based on adeno-associated virus type 1
JOURNAL      J. Virol. 73 (5), 3994-4003 (1999)
MEDLINE      99214338
PUBMED      10196295
REFERENCE      2 (bases 1 to 4718)
Xiao.W. and Wilson.J.M.
AUTHORS      Direct Submission
TITLE      Submitted (05-MAY-1998) IHGT, University of Pennsylvania, 3601
JOURNAL      Spruce Street, Philadelphia, PA 19104, USA
FEATURES      Location/Qualifiers
             1. 4718
             /organism="Adeno-associated virus 1"
             /mol_type="genomic DNA"
             /db_xref="taxon:85106"
             /note="AAV1"
             335..2206
             /codon_start=1
             /product="nonstructural protein"
             /protein_id="AAD27758.1"
             /db_xref="GI:4689098"
             /translation="MKGFEIVIKVPSDLDELHPLGIDSFVSWVAKEWELPPDSMD
             LNLQAPLVAELKQDLQWRRVSKAPEALFFVQFEGESIFHLHLILVETTVGKS
             MVLRFLSLQDLVOTIYRGIETPLNFWFATKTRNGAGGNKVVDCEYIPNLLYVKS
             TQELQWMLVLEIYSIACILAKELVAHLTHVSTQOEKNENLNPNSDAPVIRS
             KTSYRMWELMVDGITSEKQWIOQDQASYISFNAASNSRSQIKAAALDNAGKIMAL
             TKSAPDVLVGPAPPADIKTNRIIVRIELNGYEPAYAGSVFLGWAQKRGKNTIWLFG
             PATYKTNIEAANAHAVPFGCVNWTNENPFNDQVDKMWIWEHGEKMTAKVESAKA
             ILGSKVRVQDKKSSAQIDPTPIVITSNTMCVIDGNSITTFEHOQPLQDRMFKFEL
             TRLEHDFKVTQKVEHFRWAQDHVTEVAHFYVRKGGANKRPAPDADKSEPKRA
             CPSVHDPSTDAEGAPVDPADYQNKCSRHAGLMQLFPCKTCERNQNFNICFTHTG
             RDCSECFPGVSESQPVVRKTRYKLCIAIHLHLLGRAPEIACSGADLVNVDLDDCVSBQ"
             2223..4433
             /codon_start=1
             /product="capsid protein"
             /protein_id="AAD27757.1"
             /db_xref="GI:4689097"
             /translation="MAADGYLPDLNLSGIREWMDLKPAPKPKANQKQDDGRG
             LVLPGYVLGPFNGLDKGEFVNAADAAALHDKAYDQOLKAGDNPYLRLVDAEFQE
             RLOEDTSFGNLGPAVPOAKKRVLEPLGLVEGAKTAPGKR PVEOSPOEPSGIG
             KTGOQPAKRLNFGQTODSVPDPQLGEPATPAAGVTTWASGGGAPMADNNGA
             DVGNSAGNHCDSITWGDGKVIITSTRTWALPTYNHLYKQISSASTGASNDNHYFGY
             STPMGYDFNRHFCHFPDRWQRLNNNNWFRKRLNFKLFNIQVKEVTNDGVTIIA
             NNLTSTVQVFSDSYQLPYVLGSAHQCLPPFPADVMIPOYGLTLNNGSQVGRSS
             FCYLEYFSPSLRGTNNFTSYTFEEVPHSSVAHSOSLDRLMNLIDOLYLYLNRTQ
             NQSSAONKDLISRGSPAGSVQPKNLPGPCVROQVRSKTKTDNNNSNFTWTGASK
             YNLNGRSIINPGTNMAHSDKDEDKFPFMSGVMI PKESAGASNTALDNVMITDESEI
             KATNPVRSTGVANFNQSSSDTPATGDVHAMGALPGWVQDRDVLQGPFWAKLIPH
             TDGHFSPSLNGKRLGKLPPIQILKNTVPANPPAFPSATKPFASITQYSTQSVSE
             IEWLOKESGRWNPEVQVTSYNAKSANVDFTVDNNGLYTEPRPIGTRYLTREL"
BASE COUNT      1121 a 1393 c 1273 g 931 t
ORIGIN
Alignment Scores:
Pred. No.:      3,39e-226      Length:      4718
Score:          2906.00      Matches:      534
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     100.00%      Indels:      0
DB:              14      Gaps:          0
US-09-807-802a-17 (1-534) x AF063497 (1-4718)

```



```
Db 3294 CTCCTCCGTTCCCGGCGAGCTGTTTCATGATTCGCGAGTACGGCTACTCAACGCTCAAC 3353
QY 181 AenGlySerGlnAlaValGlyArgSerPheTyrCysLeuGluTyrPheProSerGln 200
Db 3354 AATGCGAGCAGGAGTGGGCGGTGTCATCTTTTACTGCTGGAATATTTCCCATCGGAG 3413
QY 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
Db 3414 ATGCTGAGAACGGCGAATTAATCTTACCTTCAGCTACCTTCGAGGAGCTGCTTTCCAC 3473
QY 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
Db 3474 AGCAGCTACGCGACAGCGAGCTGGACCGGCTGATGAATCTCTCATCGACCACTAC 3533
QY 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnIleAspLeuLeu 260
Db 3534 CTGTATTACCTGAACAGAACTCAGAACTCAGTCCGGAAGTGCACAAACAAAGGACTTGTG 3593
QY 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTyrLeuProGlyPro 280
Db 3594 TTTAGCGGGGTCTCCAGCTGGCATGCTCTGTTGACGCCCAAACTGGCTACTCGACCC 3653
QY 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
Db 3654 TGTATTCCGCGCAGCGCGCTTTAAACAAAAACAGACCAACACAGCAACTTTACC 3713
QY 301 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr 320
Db 3714 TGGACTGGTGTCTCAAAATATTAACCTTAATGGCGGTGAATCTATTAATCAACCTGGCACT 3773
QY 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
Db 3774 GCTATGGCTCTCACAAAGACGACAAAGCAAGTCTTTCCCATGAGCGGTGTCATGATT 3833
QY 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
Db 3834 TTTGGAAGAGGAGCGCGGAGCTTCAACACATGCTGCAATGGCAATGTCATGATCAGAC 3893
QY 361 GluGluGluLysLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
Db 3894 GAAGAGGAATCAAGACCACTAACCCCGTGGCCACCGAAGATTTGGGAGCTGTGGCAGTC 3953
QY 381 AsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
Db 3954 AATCTCAGGAGCAGCAGCAGACCCCTGCGACCGGAGATGTGCAATGTTATGGAGCCTTA 4013
QY 401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlnProIleTrpAlaLysIle 420
Db 4014 CCTGGAATGGTGTGGCAAGACAGACAGTATACCTGCGAGGCTCTATTTGGGCGCAAAAT 4073
QY 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
Db 4074 CCTCAGCAGGATGGACATTTTACCCGCTCTCTCTCATGGCGGCTTTGGACTTAAGCAC 4133
QY 441 ProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460
Db 4134 CGGCTCTCTCAGATCTCTCAAAAACACAGCGCTTCTCGGAATCTCTCGGAGAGTTT 4193
QY 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
Db 4194 TCGGCTACAAAGTTTGCTTCATTCATCCAGTATTTCCACAGACAGTGAAGCGGTGGAG 4253
QY 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTyrAsnProGluValGlnTyrThr 500
Db 4254 ATTGAATGGAGCTGCAGAAAGAAAACAGCAACGCTGGAATCCCCAAGTGCAGTATACA 4313
QY 501 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 520
Db 4314 TCTAATCTGCAAAATCTGCAAGTGTGATTTTCACTGTGGACAACTGAGCTTTATACT 4373
QY 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
```

```
Db 4374 GAGCTCGCCCATTTGGCACCCGTTACTCTCACCCGTCCTCCCTG 4415
RESULT 3
AF028705
LOCUS AF028705
DEFINITION Adeno-associated virus 3B, complete genome.
ACCESSION AF028705
VERSION AF028705.1 GI:2766608
KEYWORDS
SOURCE Adeno-associated virus 3B
ORGANISM Adeno-associated virus 3B
REFERENCE 1 (bases 1 to 4722)
AUTHORS Rutledge,E.A., Halbert,C.L. and Russell,D.W.
TITLE Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2
JOURNAL J. Virol. 72 (1), 309-319 (1998)
MEDLINE 98080418
PUBMED 9420229
REFERENCE 2 (bases 1 to 4722)
AUTHORS Rutledge,E.A. and Russell,D.W.
TITLE Direct Submision
JOURNAL Submitted (25-SEP-1997) Hematology, Univ of Washington, Box 357720, Seattle, WA 98195, USA
FEATURES
source
1. .4722
/organism="Adeno-associated virus 3B"
/mol_type="genomic DNA"
/db_xref="taxon:68742"
/notes="AAV3B"
317. .2191
/codon_start=1
/product="nonstructural protein Rep78"
/protein_id="AAB95451.1"
/db_xref="GI:2766608"
/translation="MPSFYEIVLVKPSDLDEHLPGISNFPVNWAEKEWELPPDSMD
PNLIEQAPLVAEKLQREFLVRRVRSAPEALFFVQFERGETTFPHLVLIETGVKS
MVVGYVSOIKELVTRIYRGEVPELPFNWFAVTRNGAGGKGVDDYDIPNLVLPK
TQPELQWANTMDQYLSACLNAERKLVAQHLTHVSTQEQONKENVNPNDAVIRS
KTASRYMELVGLVDRTGISTEKOWIQEDQASYISFNAASRSQIKALDNASIMSL
TSDAYLVGNSNPEDITKNRIYQILSLNGVDPOYASVFLGWAQKFGKRTNWLFG
PATGKTNIABIAIAHAFVFCVNTWNTENPPNDCVDKMYLWEEGKMTAKVESA
ILGGVVRVQKSKSAQIETPVTISNTNMCVADIGNSTTFPHQQPLQDMRFKEL
TRRLDHFQVTKQEVDFRFAVSDHVTDAHEFVYRGGAGKAPSNADVSBPKRQ
CTSIAQPTTSDAEPADYQNKSRHVGMLMLFPCKTCERMNQTISNVCPTHGQR
DCGCEFPQMSSEQPVSVYKXTYQKLCPIHILGRAPEIACSDLANVLDLDCVSEQ
"
2208. .4418
/codon_start=1
/product="capsid protein VP1"
/protein_id="AAB95452.1"
/db_xref="GI:2766610"
/translation="MAADGYLPDMLDNLSEGIREWALKPGVPQPKANQOQHNRKG
LVLPGYKTLGPNGLDKGEPVNEADAALAHDKAYDQQLKAGDNPYLKYHADAEFOE
RLQDTSFGNGLRAVFOAKRIILEPLGLVEEAATAPGKGRPVDOQSPBDSGSGVG
KSGQPKARKLNFQGTGDSVPDPQLGEPAPATSLGNTMSSGASNDHVPYGS
DGVGNSGNHCDSDQWLGDRTVITSTRTALPTYNHLYKQISSQSGASNDHVPYGS
TPWGVDFNPFHCFPDRWOLNNWFRPKLSFKLFINIQIKVETQNDQTTTIAN
NLTSVQVFTDSEYQLPYLGSAGQCLPPFPADVPMVPOYGLTLNNGSOAVGRSP
YCLFYFSQMLRTGNTGFPSTPDPVPHSSYAHQSILDRMLNPLIDQYLYLANTQG
TTSQTTNQSRLLFSQAGQSMGLARNMLPGCTCRQRLSKTANDNNNSNFPWTAAAK
YHLNGRSLVNPFGPAMASHKDEEKFPMHGNLIFGKEGTTASNAELDNVMTDEEI
RTTNVPATPQYGVANNLQSSNTATPTTRNDQGLPGMWQDRDVLQGTQVSWVE
TDGHPFPGLMGFLKHPPIQIMIKNTPVPANPTTSPAPKASFITQYGTQVSVSE
IEMELQKNSKRNWPEIQYTSNYSNVVDFTVDNTGVYSEPRPIGTIRYLRLN"
BASE COUNT 1204 a 1297 c 1243 g 978 t
ORIGIN
Alignment Scores:
Pred. No.: 1,266-195 Length: 4722
Score: 2530.00 Matches: 460
Percent Similarity: 91.03% Conservative: 27
Best Local Similarity: 85.98% Mismatches: 46
```

| | | | |
|---|--------|--|------|
| Query Match: | 87.06% | Indels: | 2 |
| DB: | 14 | Gaps: | 2 |
| US-09-807-802A-17 (1-534) x AF028705 (1-4722) | | | |
| QY | 1 | MetAlaSerGlyGlyGlyAlaProMetAlaAspAenAenGluGlyAlaAspGlyValGly | 20 |
| DB | 2814 | ATGGCTTCAGGCGGTGGGCGACCAATGGCGAGCAATAACGAGGGTGGCGAGTGGGT | 2873 |
| QY | 21 | AsnAlaSerGlyAsnTrpHisCysAspSerThrTrrLeuGlyAspArgValIleThr | 40 |
| DB | 2874 | AATTCCTCAGGAATTTGGCATTTGCCATTTCCCAATGGCTGGGCGACAGAGTTCATCACACC | 2933 |
| QY | 41 | SerThrArgThrTrrAlaLeuProThrTrrAenAenHisLeuTyrLysGlnIleSerSer | 60 |
| DB | 2934 | AGCACCAAGACCTGGGCGCTTGGCCACTTACACCAACCACTCTCTACACCAAACTCTCCAGC | 2993 |
| QY | 61 | AlaSerThrGlyAlaSerAenAspAenHisTyrPheGlyTyrSerThrProTrrGlyTyr | 80 |
| DB | 2994 | CAATCA---GGAGCTTCAACACGACCAACCACTACTTTGGCTACAGCACCCCTTGGGGGTAT | 3050 |
| QY | 81 | PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrrGlnArgLeuIleAen | 100 |
| DB | 3051 | TTTGACTTTAACAGATTCACATGCCACTTCTCACACGCTGACTGGCAGCACTCAATTAAC | 3110 |
| QY | 101 | AsnAenTrrGlyPheArgProLysArgLeuAenPheLysLeuPheAsnIleGlnValLys | 120 |
| DB | 3111 | AACAACTGGGGATTCGGGCCCAAGAACTCAGCTTCAGCTCTTCAACATCAAGTTAAA | 3170 |
| QY | 121 | GluValThrThrAenAspGlyValThrThrIleAlaAenAenLeuThrSerThrValGln | 140 |
| DB | 3171 | GAGGTCAAGCAGAACGATGGCAGCAGCACTATTTGCCAAATAACTTTACAGCACGGTTCAA | 3230 |
| QY | 141 | ValPheSerAspSerGluTyrGlnLeuProTrrValLeuGlySerAlaHisGlnGlyCys | 160 |
| DB | 3231 | GGTTTACGGAATTCGAGATTCAGCTCCGCTACGTGCTCGGTTCGGCGCACCAAGGCTGT | 3290 |
| QY | 161 | LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAen | 180 |
| DB | 3291 | CTCCGCGGTTTCCAGCGGACGCTTTCATGGTCCCTCAGTAGTGATACCTCACCCCTGAAC | 3350 |
| QY | 181 | AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln | 200 |
| DB | 3351 | AACGGAAGTCAACGCGTGGGACGCTCATCTTTTATCTGCTGGAGTACTTCCCTTCGCGAG | 3410 |
| QY | 201 | MetLeuArgThrGlyAsnAenPheThrPheSerTyrThrPheGluGluValProPheHis | 220 |
| DB | 3411 | ATGCTAAGGACTGGAAATTAACCTTCCAAATTCAGCTATACCTTCGAGATGTACTTTTAC | 3470 |
| QY | 221 | SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr | 240 |
| DB | 3471 | AGCAGCTACGCTCACAGCCAGAGTTTGGATCGCTTGGATGATGATGATGATGATGATGAT | 3530 |
| QY | 241 | LeuTyrTyrLeuAenArgThrGln---AenGlnSerGlySerAlaGlnAsnLysAspLeu | 259 |
| DB | 3531 | CTGTACTACTCTGAACAGACGCAAGGACCAACCTCTGGAAACCAACCAATCACCGGCTG | 3590 |
| QY | 260 | LeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAenTrrLeuProGly | 279 |
| DB | 3591 | CTTTTATGCGGCTGGGCTCAGTCTATGCTTTTGGAGCCGAGAAATTTGGCTACTCTGG | 3650 |
| QY | 280 | ProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAenSerAenPhe | 299 |
| DB | 3651 | CCCTGCTACCGGCAACAGAGACTTTCAAGAGCTGTAAACGACACCAACCAACAGTAAC | 3710 |
| QY | 300 | ThrTrrThrGlyAlaSerLysTyrAenLeuAenGlyArgGluSerIleIleAenProGly | 319 |
| DB | 3711 | CCTTGGACAGCGGCCAGCAAAATATCATCTCAATGGCGCGGACTCGCTGGTGAATCCAGA | 3770 |
| QY | 320 | ThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMet | 339 |
| DB | 3771 | CCAGCTATGCCAGTCACAAAGGACGATGAAGAAAAATTTTCCCTTATGACGCGCAATCTA | 3830 |

```

/db_xref="GI:1408468"
/translation="MPGFVEIVLVKPSDLDELPGISNSFVNVAEKSDVPPSDMD
PNLIQAPLVAKLQRELFVRRVSKAPEALFVQREKGETYFHLHLVLETIGVKS
MVGVRVQIEKLVRIYRGVEPOLPNFVTRKTRGAGGNGYVDDCYIPNLLPK
TOPELQAWTNDQVLSACLAEKRLVAQLHTVSTQEQNKQNPNSDAPVIRS
KTSARYMELGVLVDRGITSERQWIOEDQASYISFNAASNSRQIKAAALDKTMSL
KTAPDYLVSPPEDITKNRIYQILELNGDPOVAASVFLGWAOKFKGKNTIWLFG
PATTKNTAEALAHAVPFGVGVNTNENFPNDCCVDMVIMWEGKMTAKVESAKA
ILGSKVRVDCKSKSAQIEPTPIVITSNTNCAVIDGNSITFEHQQLQDRMFEEL
TRLRDHFGKTVKEQVDFRWASDRVTDVAHEFTVRKGGAKRPSADNDVSEKRE
CTSLAQPTSDAEAPADYADRYKSRHVGWMLFPCKTCERNWQISNVCFTHGQR
DCGCEFPNGMSEQPSVSVKTKYQKLCPIHTILGRAPETACSAACDLANVDLDCVSEQ
"
2209..4419

```

```

/codon_start=1
/product="capsid protein"
/protein_id="AAC5049.1"
/db_xref="GI:1408469"
/translation="WAADGYLPDLWLEDSIGIREWALKPGVPKANOQHQDNRRG
LVLPGVYKLGNGLDKGRPVNEADAALAEHDKAYDQOLKAGDNPVLKYNHADAEFOE
RLQEDTSFGNGLGRAVFOAKKRLLEPLGLVVEEAATAFGKKAQVDSQEPDPSSGVG
KGGKQARKRLNFPQGDSEVPDPOPLGEPFAAPTSLGNSNTWASGGASNDNHEGA
DVGNSGNHHCDSQWLDRVITSTRWALPTNNHLYKQISSQSGASNDNHYFGYS
TPWGYDFNRFCHFSFPRDQRLINNWGRFKCLSKFLFNTQVRGVTQNDGTTIAN
NLTSVQNFVDSYOLPYVLGAHQGLPPFADVFMVPOYGYLTLLNNGSQVGRSSF
YCLEYPPSOMLRTNNFQSYTFEDVPFHSVAHSOSLDRLANPLIDQYLYLNRTOG
TTSQNTNOSRLFPSQAGQMSLOARWLPGCYRQORLSKTDNNDNNSNFPWTAAK
YHLNGRDSLUNFGPNAWAKHDEEKFPMHGNLIFGKEGTTASNAELDNVMTIDEEI
RTNPVATQYQTVANNLQSSNAPTGTGNHOGALPGMWQDRDYLQVLPWAKVPH
TDGHPHPSGLMGFGKLPHPQIMIKTNPVNPANPTTFSPAFASFTQYSGQVSE
IEWELQKNSKWNPEIQYTSNYSNVNVDFTVDTNGVYSBRPIGTRYLTRLN"
BASE COUNT 1200 a 1295 c 1251 g 980 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 1,846-195 Length: 4726
Score: 2528.00 Matches: 459
Percent Similarity: 91.03% Conservative: 28
Best Local Similarity: 85.79% Mismatches: 46
Query Match: 86.99% Indels: 2
DB: 14 Gaps: 2

```

US-09-807-802A-17 (1-534) x AVU48704 (1-4726)

| | | | |
|----|------|--|------|
| QY | 1 | MetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGly | 20 |
| DB | 2815 | ATGGCTTCAGCGCGTGGCGCACCATGCGACACATAACAGAGGTCCGATGAGTGGGT | 2874 |
| QY | 21 | AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr | 40 |
| DB | 2875 | AATTCCTCAGAAATGGCATTCGGATTCCCAATGGCTGGCGACAGAGTCATCACACC | 2934 |
| QY | 41 | SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrIlysGlnIleSerSer | 60 |
| DB | 2935 | AGCACCAGAACCTGGGCGCTGCCACTTACAAACACCATCTCTACAGCAAAATCTCCAGC | 2994 |
| QY | 61 | AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr | 80 |
| DB | 2995 | CAATCA---GGAGCTTCAACAGCACACCATCTATTGGCTACAGCACCCCTGGGGGTAT | 3051 |
| QY | 81 | PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn | 100 |
| DB | 3052 | TTTTGACTTTAAAGATTCACACTGCCACTTCTCACCAGCTGACGTGGCAGGACCTCATTAAC | 3111 |
| QY | 101 | AsnAsnTrpGlyPheArgProIlysArgLeuAsnPhelIysLeuPheAsnIleGlnValIys | 120 |
| DB | 3112 | ARCAACTTGGGGATTCGGGCCCAAGAACTCAGCTTCAAGCTCTTCAACATCCAAAGTTAGA | 3171 |
| QY | 121 | GluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeuThrSerThrValGln | 140 |
| DB | 3172 | GGGGTCAGCAGAACGATGCGACGACGACTATTGGCCCAATTAACCTTACCAGCAGCGTTCAA | 3231 |
| QY | 141 | ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys | 160 |
| DB | 3232 | GTGTTTACGGAGCTCGAGTATCAGCTCCGCTACGCTCGGTTCGGGCGCACCAAGGCTGT | 3291 |
| QY | 161 | LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn | 180 |
| DB | 3292 | CTCCCGCGGTTCAGCGGGAGCTTCTCATGTCCTCTAGTAGTATGATACCTCACCCGTGAAC | 3351 |
| QY | 181 | AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln | 200 |
| DB | 3352 | AACGGAAGTCAAGCGGTGGGAGCGCTCATCTTTTACTGCTGGAGTACTTCCCTTCGAG | 3411 |
| QY | 201 | MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis | 220 |
| DB | 3412 | ATGCTAAGGACTGGAAATAAATCTCAATTCAGCTATACCTTCGAGGATGTACCTTTTCAC | 3471 |
| QY | 221 | SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr | 240 |
| DB | 3472 | AGCAGTACGCTCAGCAGCAGATTTGGATTCGCTTGGATTCCTCTTATTGATCAGTAT | 3531 |
| QY | 241 | LeuTyrTyrLeuAsnArgThrGln---AsnGlnSerGlySerAlaGlnAsnLysAspLeu | 259 |
| DB | 3532 | CTGTACTACTGACAGAACAGACCAAGCAACACCTCTGGAAACAACCAACCAATCACGGCTG | 3591 |
| QY | 260 | LeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGly | 279 |
| DB | 3592 | CTTTTTCAGCAGGCTGGGCTCAGTCTATGCTTTGACAGCGCAGAAATTTGGCTACCTGG | 3651 |
| QY | 280 | ProCysTyrArgGlnArgValSerIlySerIlyThrAspAsnAsnAsnSerAsnPhe | 299 |
| DB | 3652 | CCCTGCTACCGGCAACAGAGACTTTCAAGAGCTGTAAACGACCAACACACAGTAGTAATTT | 3711 |
| QY | 300 | ThrTrpThrGlyAlaSerIlyTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGly | 319 |
| DB | 3712 | CCTTGGACAGCGGCGCAGCAATATCATCTCAATGGCGCGACTCGCTGTGTAATCAGGA | 3771 |
| QY | 320 | ThrAlaMetAlaSerHisIlyAspAspGluAspLysPhePheProMetSerGlyValMet | 339 |
| DB | 3772 | CCAGCTATGGCGAGTCACAGAGACGATGAAGAAAATTTTCCCTTATGCACGCGCAATCTA | 3831 |
| QY | 340 | IlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThr | 359 |
| DB | 3832 | ATATTGGCAAGAAAGAGGACAAACGCAAGTAGTAACGAGAAATAGATTAATGATTAGC | 3891 |
| QY | 360 | AspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAla | 379 |
| DB | 3892 | GATGAAGAAGAGATTCGTACCAACCAATCTCTGTGGCAACAGAGCAGTATGGAATCTGGCA | 3951 |
| QY | 380 | ValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAla | 399 |
| DB | 3952 | AATAACTTTCAGAGCTCAATACAGCTCCCAAGCTGGAACCTGTCATCATCAGGGGGCC | 4011 |
| QY | 400 | LeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLys | 419 |
| DB | 4012 | TTACTTGGCATGCTGTGGCAAGATCTGACGTGTACCTTCAAGGACCTATCTTGGGCAAG | 4071 |
| QY | 420 | IleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLys | 439 |
| DB | 4072 | ATTCCTCAGCAGGATGACACTTTTCATCTCTCTGATGGAGGCTTTGAGCTGAA | 4131 |
| QY | 440 | AsnProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGlu | 459 |
| DB | 4132 | CATCCGCTCTCTCAATCATGATCAAAAATACTCCGGTACCGGCAATCTCCCGAGACT | 4191 |
| QY | 460 | PheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerVal | 479 |
| DB | 4192 | TTACGCGCGCGCCAGTTTGGCTTCATTTATCTCAGTACTCTCAGTGGACAGGTTCAGCGT | 4251 |
| QY | 480 | GluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyr | 499 |
| DB | 4252 | GAATTTGAGTGGAGCTACAGAAAGAAACAGCAACGTTGGAATCCAGAGATTCAGTAC | 4311 |
| QY | 500 | ThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyr | 519 |

Db 4312 ACTTCCAACTACAAAGTCTGTTAATGTGGACTTACTGTAGACATAATGGTGTAT 4371
Qy 520 ThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
Db 4372 AGTGAACCTCGCCCTATTGGAAACCGGTATCTCACACGAACTTG 4416

RESULT 5

AR264580
LOCUS AR264580 8179 bp DNA linear PAT 10-APR-2003
DEFINITION Sequence 5 from patent US 6491907.
ACCESSION AR264580
VERSION AR264580.1 GI:29692850
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

REFERENCE

1 (bases 1 to 8179)
Rabinowitz, J.E., Samulski, R.J. and Xiao, W.
AUTHORS Recombinant parvovirus vectors and method of making
TITLE Patent: US 6491907-A 5 10-DEC-2002;
JOURNAL Location/Qualifiers
FEATURES
1. 8179
/organism="unknown"

BASE COUNT 2082 a 2078 c 2027 g 1992 t
ORIGIN

Alignment Scores:

Pred. No.: 2,47e-194 Length: 8179
Score: 2518.00 Matches: 458
Percent Similarity: 90.65% Conservative: 27
Best Local Similarity: 85.61% Mismatches: 48
Query Match: 86.65% Indels: 2
Gaps: 2

US-09-807-802A-17 (1-534) x AR264580 (1-8179)

Qy 1 MetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGly 20
Db 2739 ATGGCTACAGGCGAGTGGCGCCCAATGGCAGACAATAACGAGGGCGCGAGTGGGT 2798
Qy 21 AsnAlaSerGlyAsnTyrHisCysAspSerThrTyrLeuGlyAspArgValIleThr 40
Db 2799 AATTCCTCGGAAATTTGGCATTTGCCATTCACATGGATGGCGCAGAGTTCATCACACC 2858
Qy 41 SerThrArgThrTyrAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSer 60
Db 2859 AGCACCAGACTGGGCGCTGCCACCTACACACACACCTCTACAAACAATTTCCAGC 2918
Qy 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTyrGlyTyr 80
Db 2919 CAATCA--GGAGCCTCGAACGACAATCACTACTTTGGCTACAGCACCCCTTGGGGGTAT 2975
Qy 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTyrGlnArgLeuIleAsn 100
Db 2976 TTGACTTCAACAGATTCCAGTGGCCACTTTCCACCGCTGACTGGCAAGACTCATCAAC 3035
Qy 101 AsnAsnTyrGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 120
Db 3036 AACAACTGGGATTCGGACCCCAAGAGACTCAACTTCAAGCTTTTAAACATCAAGTCAA 3095
Qy 121 GluValThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 140
Db 3096 GAGGTTCAGCAGATGACGGTACGACGACGAGATTGGCAATTAACCTTACCAGCAGCGTTCCAG 3155
Qy 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
Db 3156 GTGTTTACTGACTCCGAGTACCAGCTCCCGTACGCTCGGTGCTCGGTCGCGCACCAAGCGTGT 3215
Qy 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
Db 3216 CTCGCCCGCTTTCAGAGGAGCGTCTTTCATGGTCCCTCAGTATGGATACCTCACCCCTGAC 3275
Qy 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200

Db 3276 AACGGAAGTCAAGCGGTGGGACGCTCATCTTTTACTGCCTGGAGTACTTCCCTTGGCAG 3335
Qy 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
Db 3336 ATGCTAAGGACTGGAAATAACTTTCCAATTCAGCTATATACCTTCGAGGATGTATACCTTTTCAC 3395
Qy 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
Db 3396 AGCAGCTACGCTCACGCCAGAGTTTGGATCGCTTGATGAATCTCTTATTGATCAGTAT 3455
Qy 241 LeuTyrTyrLeuAsnArgThrGln--AsnGlnSerGlySerAlaGlnAsnLysAspLeu 259
Db 3456 CTGTACTACTCTGAACAGACGCAAGCAACACCTCTCGAACCAACCAACCAATCATCCGCGTG 3515
Qy 260 LeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTyrLeuProGly 279
Db 3516 CTTTTTAGCCAGCGTGGCGCTCAGCTATGTCTTTGACGCGCCAGAAATTTGGCTACCTGGG 3575
Qy 280 ProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPhe 299
Db 3576 CCTGTCTACCGGCAACAGAGACTTTCAAGACTCTTAACGACGACCAACACACAGTAACCTTT 3635
Qy 300 ThrTyrThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGly 319
Db 3636 CCTTGGACAGCGCCAGCAAAATATCATCTCAATGGCGCGACTCGCTGGTGAATCCAGGA 3695
Qy 320 ThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMet 339
Db 3696 CCAGCTATGGCCAGTCAAGGACGATGAAGAAATTTTTCCTATGACGCGCAATCTA 3755
Qy 340 IlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThr 359
Db 3756 ATATTTGCAAGAGAGGACCAACCGCAAGTAAACGACAGATTTAGATAATGTAATGATTACG 3815
Qy 360 AspGluGluIleLysAlaThrAsnProValIleThrGluArgPheGlyThrValAla 379
Db 3816 GATCAAGAAGAGATTCGTACCCCAATCCTGTGCAACAGACAGTATGGAACCTGTGCA 3875
Qy 380 ValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyVala 399
Db 3876 AATAACTTGCAGAGCTCAATATACAGCTCCACGACTGGAACTGCAATCATCAGGGGGCC 3935
Qy 400 LeuProGlyMetValTyrGlnAspArgAspValTyrLeuGlnGlyProIleTyrAlaLys 419
Db 3936 TTACTCTGGCATGTGTGGCAAGATCGTGACGTGTACCTTCAAGGACCTATCTGGGCAAG 3995
Qy 420 IleProHisThrAspGlyHisPheHisPheSerProLeuMetGlyGlyPheGlyLeuLys 439
Db 3996 ATTCTTCACACGGATGGACACTTTTCATCTCTCTCTGATGGGAGGCTTTGGACTGAA 4055
Qy 440 AsnProProGlnIleLeuIleLysAsnThrProValProAlaAsnProAlaGlu 459
Db 4056 CTTCCGCTCTCTCAATCATGATCAAAATACTCCGGTACTCGAATCTCTTCAGCCACC 4115
Qy 460 PheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerVal 479
Db 4116 TTCAGTGGGCAAGTTTGTCTTCTTCATCACACAGTACTCCACGGGACAGGTTCAGCGTG 4175
Qy 480 GluIleGluTyrGluLeuGlnLysGluAsnSerLysArgTyrAsnProGluValGlnTyr 499
Db 4176 GAGATCGAGTGGGAGCTGCAGAGAGAAACAGCAACGCTGGAATCCCGAATTCAGTAC 4235
Qy 500 ThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyr 519
Db 4236 ACTTCCAACTACAAAGTCTGTTAATCGTGGACTTACCGTGGATACTAATGGCGGTAT 4295
Qy 520 ThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
Db 4296 TCAGAGCTCGCCCATTTGGCACCAGATACCTGACTCGTAACTCTG 4340

RESULT 6
AR222046

| | | | | | |
|-----------------|---|--|------|--|-----------------------------------|
| D | b | | 2747 | AGCAGCTACGCTCAAGCCAGAGTCTGGACCGTCTCATGAATCTCTCATCGACCAGTAC | 2806 |
| Q | y | | 241 | LeuTyTrpLeuAenArghThrGlnAeNgInSerGlySerAlaGlnAenLysAspLeuLeu | 260 |
| D | b | | 2807 | CTGTATTACTTGAGCAGAAACAACACTCCAAGTGGAACACCAACGAGTCNAAGGCTTCAG | 2866 |
| Q | y | | 261 | PheSerArgGlySerProAlaGlyMetSerValGlnProLysAenTrpLeuProGlyPro | 280 |
| D | b | | 2867 | TTTTTCTCAGCCGGAGGAGTGCATTTCCGGACCAAGTCTAGGAAGTCTCTCGTGACCC | 2926 |
| Q | y | | 281 | CyeTyArgGlnGlnArgValSerLysThrLysPheAenAenAenSerAsnPheThr | 300 |
| D | b | | 2927 | TGTTACCGCCAGCAGCGAGTATCAAAGACATCTCGGATTAACAACAACAGTAATATCG | 2986 |
| Q | y | | 301 | TriThrClalSerLysTyrrAnLeuAenGlyVargLuSerIlelleAenProGlyThr | 320 |
| D | b | | 2987 | TGAGCTGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTGGTGAATCCGGGCCG | 3046 |
| Q | y | | 321 | AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle | 340 |
| D | b | | 3047 | GCCATGGCAAGCCACAAGGACGATGAAGAAAAGTTTTTCTCAGAGCGGGTTCTCATC | 3106 |
| Q | y | | 341 | PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAenAenValMetIleThrAsp | 360 |
| D | b | | 3107 | TTTTGGGAAGCAAGGCTCAGAGAAAAACAATGTGGACATTTGAAAAGGTCATGATTACAGAC | 3166 |
| Q | y | | 361 | GluGluGluIleLysAlaThrAenProValAlaThrGluAargPheGlyThrValAlaVal | 380 |
| D | b | | 3167 | GAAGAGGAATCAGGACACCAATCCCGTGGCTACGGAGCAGTATGGTCTGTATCTACC | 3226 |
| Q | y | | 381 | AsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu | 400 |
| D | b | | 3227 | AACCTCCAGAGAGGCAACAGCACAGCAGCTACCGCAGATGTCAACACACAAGGCGTCTT | 3286 |
| Q | y | | 401 | ProGlyMetValTrpGlnAspArgAspValTyrrLeuGlnGlyProIleTrpAlaLysIle | 420 |
| D | b | | 3287 | CCAGGCATGGTCTGGCAGGACAGATGTGTACCTTCAGGGGCCCATCTGGSCAAGAANT | 3346 |
| Q | y | | 421 | ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAen | 440 |
| D | b | | 3347 | CCACACAGCGAGCGACATTTTCCACCCCTCTCCCCTCATGGGTGATTCGGACTTAAACAC | 3406 |
| Q | y | | 441 | ProProProGlnIleLeuIleLysAenThrProValProAlaAanProProAlaGluPhe | 460 |
| D | b | | 3407 | CCTCTCTCCAGATTTCTCATCAAGAACACCCCCGGTACTCTCGGAATCTCTCGACCACTTC | 3466 |
| Q | y | | 461 | SerAlaThrLysPheAlaSerPheIleThrGlnTyrrSerThrGlyGlnValSerValGlu | 480 |
| D | b | | 3467 | AGTGGCGCAAGTTTGTCTCTTCATCACACAGTACTCCAGCGACAGGTCAGCGTGAG | 3526 |
| Q | y | | 481 | IleGluTrpGluLeuGlnLysGluAenSerLysArgTrpAanProGluValGlnTyrrThr | 500 |
| D | b | | 3527 | ATCAGTGGGAGCTGTCAGAGAGAAAAACAGCAACGCTGGAAATCCCGAAATTCAGGTACACT | 3586 |
| Q | y | | 501 | SerAenTyrrAlaLysSerAlaAenValAspPheThrValAspAenAenGlyLeuTyrrThr | 520 |
| D | b | | 3587 | TCCAACTACAAACAGTCTGTAAATGTGGACTTTACTGTGGACACTAATGGCGTGATTCA | 3646 |
| Q | y | | 521 | GluProArgProIleGlyThrArgTyrrLeuThrArgProLeu | 534 |
| D | b | | 3647 | GAGCCTCGCCCAATGGCACCAGATACCTGACTCGTAATCTG | 3688 |
| RESULT 7 | | | | | |
| AX205074 | | | | | |
| LOCUS | | | | | 4072 bp |
| DEFINITION | | | | | DNA linear PAT 30-AUG-2000 |
| ACCESSION | | | | | Sequence 4 from Patent WO0155361. |
| VERSION | | | | | AX205074.1 GI:15394315 |
| KEYWORDS | | | | | unidentified |
| SOURCE | | | | | unidentified |
| ORGANISM | | | | | unclassified |

REFERENCE

1 Hardy, S.F.
 Recombinant aav packaging systems
 TITLE Patent: WO 0155361-A 4 02-AUG-2001;
 JOURNAL Chiron Corporation (US)
 FEATURES Location/Qualifiers
 source 1..4072

/organism="unidentified"
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"
 /note="recombinant DNA"

BASE COUNT 1073 a 1025 c 992 g 982 t
 ORIGIN

Alignment Scores:

Pred. No.: 9,09e-192 Length: 4072
 Score: 2481.50 Matches: 445
 Percent Similarity: 90.26% Conservative: 37
 Best Local Similarity: 83.33% Mismatches: 51
 Query Match: 85.39% Indels: 1
 DB: 6 Gaps: 1

US-09-807-802A-17 (1-534) x AX205074 (1-4072)

QY 1 MetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGly 20
 DB 2090 ATGGCTACAGGAGTGGCGCACAATGGCAGACAAATAGGAGGCGCGGAGTGGGT 2149
 QY 21 AsnAlaSerGlyAsnTrpHisCysAspSerTrpLeuGlyAspArgValIleThr 40
 DB 2150 AATTCCTCGGAAATGGCATTTGGATTCACATGGATGGGCGACAGAGTCATCACACC 2209
 QY 41 SerThrArgThrTrpAlaLeuProThrTyraAsnHisLeuTyLysGlnIleSerSer 60
 DB 2210 AGCACCCGAACTGGGCGCTGCCACCTACAAACACCACTCTACAAACAAATTTCCAGC 2269
 QY 61 AlaSerThrGlyAlaSerAsnAspHisTyPheGlyTySerThrProTrpGlyTy 80
 DB 2270 CAATCA---GGAGCTCGAACACCAATCACTACTTTGGCTACAGCACCCCTTGGGGGTAT 2326
 QY 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
 DB 2327 TTTGACTTCAACAGATTCCACTGCCACTTTTCCACACGTGACTGGCAAGACTCATCAAC 2386
 QY 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 120
 DB 2387 AACAACTGGGGATTCGACCCCAAGAGACTCAACTTCAAGCTCTTTAAACATTCAAGTCAA 2446
 QY 121 GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 140
 DB 2447 GAGGTCACGCAGAAATGACGGTACGACGAGATTGCCAATAACCTTTACGACGCGTTTCCAG 2506
 QY 141 ValPheSerAspSerGluTyrglnLeuProTyValLeuGlySerAlaHisGlnGlyCys 160
 DB 2507 GTGTTTACTGACTCGGAGTACAGCTCCGTACGTCTCGGCTCGGCGCATCAAGATGC 2566
 QY 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrglyTyLeuThrLeuAsn 180
 DB 2567 CTCGCGCGCTTCCAGCAGACGCTCTCATGTGTGCACAGTATGGATACCTCACCCCTGAAC 2626
 QY 181 AsnGlySerGlnAlaValGlyArgSerPheTyrcysLeuGluTyPheProSerGln 200
 DB 2627 AACGGGAGTCAGGAGTAGGACGCTCTTCATTTTACTGCTGGAGTACTTCTCTTCTCAG 2686
 QY 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyThrPheGluGluValProPheHis 220
 DB 2687 ATGCTGGTACCGGAAACAACTTTTACCTTACCTACCTACCTTTTGAGGAGCTTCTTCCAC 2746
 QY 221 SerSerTyAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTy 240
 DB 2747 AGCAGCTACGCTCACAGCCAGAGTCTGGACCGCTCTCATGAATCTCTCATCGACCACTAC 2806
 QY 241 LeuTyTyTyLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260

Db 2807 CTGTATTACTTGGACAGAACAAACACTCCAAGTGGAAACCAACGACAGTCAAGGCTTCAG 2866
 QY 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
 Db 2867 TTTTCTCAGCGCGAGCGAGTGCATCTCGGACAGCTCTAGAACTGGCTCTCTGGGACC 2926
 QY 281 CysTyArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
 Db 2927 TGTTCACGCCAGCAGCAGATATCAAGACATCTCGGATTAACAACAACAGTGAATACTCG 2986
 QY 301 TrpThrGlyAlaSerLysTyraAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr 320
 Db 2987 TGGACTGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTGGTGAATCCGGGCGCG 3046
 QY 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
 Db 3047 GCCATGGCAGCCACAGGAGCATGAAGAAAGTTTTCCTCAGAGCGGGTCTCATC 3106
 QY 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
 Db 3107 TTTGGGAAGCAAGGCTCAGAGAAAAAATAATGTGACATTTGAAAAGGTCTATGATTACAGAC 3166
 QY 361 GluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
 Db 3167 GAAGAGGAATCAGGACAACCAATCCGTGGCTACGGAGCAGTATGGTCTGTATCTACC 3226
 QY 381 AsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
 Db 3227 AACCTCCAGAGAGCAACAGACAGACTACCCAGATGTCAACACACAGGCGTTCTT 3286
 QY 401 ProGlyMetValTrpGlnAspArgAspValTyLeuGlnGlyProIleTrpAlaLysIle 420
 Db 3287 CCAGGCATGGTCTGGCAGGACAGAGATGTGTACCTTCAGGGGCGCCATCTGGGCAAGATT 3346
 QY 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
 Db 3347 CCACACGCGAGCGACATTTTCACCCCTCTCCCTCATGGGTGGATTCCGACTTTAAACAC 3406
 QY 441 ProProGlnIleLeuLysAsnThrProValProAlaAsnProProAlaGluPhe 460
 Db 3407 CCTCTCCACAGATTCTCATCAAGAACACCCCGTACTTGGGAATCTTCCAGCACCTTC 3466
 QY 461 SerAlaThrLysPheAlaSerPheIleThrGlnTySerThrGlyGlnValSerValGlu 480
 Db 3467 AGTGGCGCAAGTTTGTCTTCTTATCACAAGTACTCCACGGGACAGGTCAAGCGTGGAG 3526
 QY 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyThr 500
 Db 3527 ATCAGTGGGAGCTGCGAAGAGAAACAGCAAAACGCTGGAAATCCGAAATTCAGTACT 3586
 QY 501 SerAsnTyAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyThr 520
 Db 3587 TCCAACTACAACAAGTCTGTTAATGTGACTTTTACTGTGGACACTAATGGCGTGTATTCA 3646
 QY 521 GluProArgProIleGlyThrArgTyLeuThrArgProLeu 534
 Db 3647 GAGCTCGCCCATTTGGCACCAGATACCTGACTCGTAATCTG 3688

RESULT 8

AX282480

LOCUS

DEFINITION

AX282480

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AX282480 4679 bp DNA linear PAT 02-NOV-2001
 Sequence 1 from Patent WO016888.

AX282480

AX282480.1 GI:16609611

Adeno-associated virus 2

Adeno-associated virus 2

Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.

Xiao, W. and Daring, M.J.

Production of chimeric capsid vectors

Patent: WO 016888-A 1 20-SEP-2001;

Neurologix, Inc. (US)
 Location/Qualifiers
 1. 4679
 /organism="Adeno-associated virus 2"
 /mol_type="genomic DNA"
 /db_xref="taxon:10804"
 BASE COUNT 1198 a 1262 c 1255 g 964 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1,09e-191 Length: 4679
 Score: 2481.50 Matches: 445
 Percent Similarity: 90.26% Conservations: 37
 Best Local Similarity: 83.33% Mismatches: 51
 Query Match: 85.39% Indels: 1
 DB: 6 Gaps: 1

US-09-807-802A-17 (1-534) x AX282480 (1-4679)

QY 1 MetAlaSerGlyClyAlaProMetAlaAspAsnGluGlyAlaAspGlyValGly 20
 DB 2809 ATGGCTACAGGAGTGGCGCACCACCAATGGCAGACAAATACGAGGGCGCCGACGGAGTGGGT 2868

QY 21 AsnAlaSerGlyValAsnTrpHisCysAspSerThrTyrLeuGlyValAspArgValIleThr 40
 DB 2869 AATTCCTCGGAAATTTGGCAATTCGGATTCACATGATGGCGCAGAGTCAATCACCACC 2928

QY 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrIleGlnSerSer 60
 DB 2929 AGCACCAGAACCTGGGCGCTCCGCCACCTACCAACACCACTCTACCAACAAATTTCCAGC 2988

QY 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
 DB 2989 CAATCA---GGAGCCCTCGAACGCAATCACTACTTTGGCTACAGCACCCCTTGGGGGTAT 3045

QY 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
 DB 3046 TTTGACTTCAACAGATTCACCTGCCACTTTTACCAGGTGACTGGCAAGAGTCAATCAAC 3105

QY 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 120
 DB 3106 AACAACTGGGGATTCGACCCCAAGAGACTCAACTTCAAGCTCTTTAACTTCAAGTCAAA 3165

QY 121 GluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeuThrSerThrValGln 140
 DB 3166 GAGTCAAGCAGAAATGACGTATACGACGACGATTCGCAATTAACCTTACCAGCAGGTTCAG 3225

QY 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
 DB 3226 GTGTTTACTGACTCGGAGTACAGCTCCCTACGTCTCGCTCGGCGCATCAAGGATGC 3285

QY 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
 DB 3286 CTCGCGCGGTTCCAGCAGACAGCTCTTATGGTGCCACAGATGATGATGATCACTCAACCTGAAC 3345

QY 181 AsnGlySerGlnAlaValArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
 DB 3346 AACGGAGTCAGGAGTAGGACGCTCTTCATTTTACTGCTCGGAGTACTTTCCTTCTCAG 3405

QY 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
 DB 3406 ATGCTGCTACCGGAACAACTTTACCTTCAGCTACACTTTTGGAGGAGCTTCTTCCAC 3465

QY 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
 DB 3466 AGCAGCTACGCTCACAGCCAGAGTCTGGACCGTCTCATGAATCTCTCATCGCACCAGTAC 3525

QY 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
 DB 3526 CTGTATTACTTGGCAGAACAAACACCTCCAGTGGAAACCCACGCGAGTCAAGGCTTCAG 3585

QY 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
 DB 3585

DB 3586 TTTTCTCAGCGCGGAGCGAGTGACATTCGGGACGAGCTCTAGGAACCTGGCTTCTCTGGACCC 3645

QY 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
 DB 3646 TGTACCGCCAGCAGCAGGATATCAAGACATCTCGGATAACAACAACAGTGAATACTCG 3705

QY 301 TrpThrGlyAlaSerLysThrAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr 320
 DB 3706 TGGACTGGAGCTACCAAGTACCCTCAATGGCAGAGACTCTCTGGTGAATCCGGGCGCCG 3765

QY 321 AlaMetAlaSerHisLysAspGluAspLysPhePheProMetSerGlyValMetIle 340
 DB 3766 GCCATGGCAAGCCACAGGACGATGAAGAAAGTTTTTCTCAGAGCGGGTCTCATC 3825

QY 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
 DB 3826 TTTGGAGCAAGGCTCAGAGAAACAATGTGACATTTGAAGAAGGTTCATGATTACACAC 3885

QY 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
 DB 3886 GAAGAGGAATCAGGACACCAATCCCTGGCTACGAGCAGATGGTCTCTGTATCTACC 3945

QY 381 AsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
 DB 3946 AACCTCCAGAGAGCGCAACAGACAGCTACCGCAGATGTCAACACACAAGGCGTCTT 4005

QY 401 ProGlyMetValTrpGlnAspArgValTyrLeuGlnGlyProIleTrpAlaLysIle 420
 DB 4006 CAGGATGGTCTGGCAGGACAGAGTGTACCTTTCAGGGGCGCCATCTGGGCAAGATT 4065

QY 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
 DB 4066 CCACACAGGAGCGACATTTTCAACCCCTCTCCCTCATGGGTGGATTCGGACTTAAACAC 4125

QY 441 ProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460
 DB 4126 CTCCTCCACAGATTTCTCATCAAGAACACCCCGGTACCTGGGAATCTTTCGACCACTTC 4185

QY 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
 DB 4186 AGTGGCGCAAGTTTGTCTTCTTCATCACAGTACTCACGGACAGGTCAAGCTGGAG 4245

QY 481 IleGluTyrGluLeuGlnLysGluAsnSerLysArgTyrAsnProGluValGlnTyrThr 500
 DB 4246 ATCGAGTGGAGTCTCAGAGAAACAGCAACCTGGAATCCCGAAATTCAGTACACT 4305

QY 501 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 520
 DB 4306 TCCTCACTACACAGCTCTGTTAATGTGACTTTTACTGTGGACACTAATGGCGGTATTCA 4365

QY 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
 DB 4366 GAGCCTCGCCCATTTGGCACCAGATACCTGACTCGTAATCTG 4407

RESULT 9
 AF043303 4679 bp DNA linear VRL 24-FEB-1998
 LOCUS Adeno-associated virus 2, complete genome.
 DEFINITION AF043303
 ACCESSION AF043303.1 GI:2906016
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Adeno-associated virus 2
 Adeno-associated virus 2
 Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
 Ruffing, M., Heid, H. and Kleinschmidt, J. A.
 Mutations in the carboxy terminus of adeno-associated virus 2
 capsid proteins affect viral infectivity: lack of an RGD
 integrin-binding motif
 J. Gen. Virol. 75 (Pt 12), 3385-3392 (1994)
 MEDLINE 9508582
 PUBMED 7996133
 REFERENCE 2 (bases 1 to 4679)

AUTHORS Berns, K.I., Bohenzky, R.A., Cassinotti, P., Colvin, D., Donahue, B.A., Dull, T., Horer, M., Kleinschmidt, J.A., Ruffing, M., Snyder, R.O., Tratschin, J.-D. and Weitz, M.

TITLE

Direct Submission

JOURNAL Submitted (15-JAN-1998) Cell Genesys Inc., 342 Lakeside Dr., Foster City, CA 94404, USA

FEATURES

Location/Qualifiers

source

1..4679

/organism="Adeno-associated virus 2"

/mol_type="genomic DNA"

/db_xref="taxon:10804"

/note="changes relative to the original sequence, GenBank Accession Number J01901, have been detected and verified by several different laboratories"

repeat_region

1..145

/note="inverted terminal repeat"

/rpt_type="inverted"

42..83

misc_feature

287..4451

/note="flip oriented DNA"

precursor_RNA

Join(321..1906,2228..2252)

CDS

/codon_start=1

/product="Rep 68 protein"

/protein_id="AAC03774.1"

/db_xref="GI:2906017"

/translation="MPGFEYIVIKVPSDLDEHLPGISDSFVNWVAEKEMWELPPDSMD LNLBQALTVAEKLQRLTEWRVSKAPEALFVQPEKGSYFHHMLVETTVGVS MVLGRFSQIRKLIQRIYRGIEPTLPNFAVTKTRNGAGGKQVDECIYIPNLLPK TQPELQWANTMEQYLSACLNLTERKLVAQHLTHVSQTQONKENQNPNSDAVIRS KTSARYMELVNDKIGTSEKQWIOEQDQASYISFNAASNSRSQIKAAALDQAKIMSL TKTADPYLVQGPVEDISNNRIYKILELNGYDPOYAAVSFLGWATKFKGKNTIWLFG PATTGKTNIAEIAHTVPFYGCYVWNTNENFPNDCDQVNIWEEGKMTAKVSESAKA ILGGSKVRDQKCKSSAQIDPTPVIIVTSNTNNCAVIDGNTTFEHOQLQDRMFKFEL TRRLDHDGKVTQKQVDFFRWAKDHVVEHEFVYKGGAKKRPAPSDADISBPKRV RESVAQSTSDAESINADRLARGHSL"

321..2186

CDS

/codon_start=1

/product="Rep 78 protein"

/protein_id="AAC03775.1"

/db_xref="GI:2906018"

/translation="MPGFEYIVIKVPSDLDEHLPGISDSFVNWVAEKEMWELPPDSMD LNLBQALTVAEKLQRLTEWRVSKAPEALFVQPEKGSYFHHMLVETTVGVS MVLGRFSQIRKLIQRIYRGIEPTLPNFAVTKTRNGAGGKQVDECIYIPNLLPK TQPELQWANTMEQYLSACLNLTERKLVAQHLTHVSQTQONKENQNPNSDAVIRS KTSARYMELVNDKIGTSEKQWIOEQDQASYISFNAASNSRSQIKAAALDQAKIMSL TKTADPYLVQGPVEDISNNRIYKILELNGYDPOYAAVSFLGWATKFKGKNTIWLFG PATTGKTNIAEIAHTVPFYGCYVWNTNENFPNDCDQVNIWEEGKMTAKVSESAKA ILGGSKVRDQKCKSSAQIDPTPVIIVTSNTNNCAVIDGNTTFEHOQLQDRMFKFEL TRRLDHDGKVTQKQVDFFRWAKDHVVEHEFVYKGGAKKRPAPSDADISBPKRV RESVAQSTSDAESINADRLARGHSL"

CDS

321..2186

/codon_start=1

/product="Rep 78 protein"

/protein_id="AAC03775.1"

/db_xref="GI:2906018"

/translation="MPGFEYIVIKVPSDLDEHLPGISDSFVNWVAEKEMWELPPDSMD LNLBQALTVAEKLQRLTEWRVSKAPEALFVQPEKGSYFHHMLVETTVGVS MVLGRFSQIRKLIQRIYRGIEPTLPNFAVTKTRNGAGGKQVDECIYIPNLLPK TQPELQWANTMEQYLSACLNLTERKLVAQHLTHVSQTQONKENQNPNSDAVIRS KTSARYMELVNDKIGTSEKQWIOEQDQASYISFNAASNSRSQIKAAALDQAKIMSL TKTADPYLVQGPVEDISNNRIYKILELNGYDPOYAAVSFLGWATKFKGKNTIWLFG PATTGKTNIAEIAHTVPFYGCYVWNTNENFPNDCDQVNIWEEGKMTAKVSESAKA ILGGSKVRDQKCKSSAQIDPTPVIIVTSNTNNCAVIDGNTTFEHOQLQDRMFKFEL TRRLDHDGKVTQKQVDFFRWAKDHVVEHEFVYKGGAKKRPAPSDADISBPKRV RESVAQSTSDAESINADRYONKSRHGVNMLFPFCQCRMNQNSNCTFTHGQK DCLECFPVSESQPVSVVKKAYQKLCYIHHIMGVKVPDCTACDLVNVLDLDDCIFEQ"

370

variation

/note="compared to sequence of GenBank Accession Number J01901"

/replace="g"

878..4451

Join(993..1906,2228..2252)

/codon_start=1

/product="Rep 40 protein"

/protein_id="AAC03776.1"

/db_xref="GI:2906019"

/translation="MELVGLVLDKGITSEKOWIQBDQASYISFNAASNSRSQIKAAALD RNTILWGPATTKTNIAEIAHTVPFYGCYVWNTNENFPNDCDQVNIWEEGKMTA KVESAKILGSKVRDQKCKSSAQIDPTPVIIVTSNTNNCAVIDGNTTFEHOQLQ DRMFPELTRLDHDFGKVTQKQVDFFRWAKDHVVEHEFVYKGGAKKRPAPSDA DISEPKRVRESVAQSTSDAESINADRLARGHSL"

993..2186

CDS

/codon_start=1

/product="Rep 52 protein"

/protein_id="AAC03777.1"

/db_xref="GI:2906020"

/translation="MELVGLVLDKGITSEKOWIQBDQASYISFNAASNSRSQIKAAALD RNTILWGPATTKTNIAEIAHTVPFYGCYVWNTNENFPNDCDQVNIWEEGKMTA KVESAKILGSKVRDQKCKSSAQIDPTPVIIVTSNTNNCAVIDGNTTFEHOQLQ DRMFPELTRLDHDFGKVTQKQVDFFRWAKDHVVEHEFVYKGGAKKRPAPSDA DISEPKRVRESVAQSTSDAESINADRYONKSRHGVNMLFPFCQCRMNQNSNCTFTHGQK DCLECFPVSESQPVSVVKKAYQKLCYIHHIMGVKVPDCTACDLVNVLDLDDCIFEQ"

370

variation

/note="compared to sequence of GenBank Accession Number J01901"

/replace="g"

878..4451

Join(993..1906,2228..2252)

/codon_start=1

/product="Rep 40 protein"

/protein_id="AAC03776.1"

/db_xref="GI:2906019"

/translation="MELVGLVLDKGITSEKOWIQBDQASYISFNAASNSRSQIKAAALD RNTILWGPATTKTNIAEIAHTVPFYGCYVWNTNENFPNDCDQVNIWEEGKMTA KVESAKILGSKVRDQKCKSSAQIDPTPVIIVTSNTNNCAVIDGNTTFEHOQLQ DRMFPELTRLDHDFGKVTQKQVDFFRWAKDHVVEHEFVYKGGAKKRPAPSDA DISEPKRVRESVAQSTSDAESINADRYONKSRHGVNMLFPFCQCRMNQNSNCTFTHGQK DCLECFPVSESQPVSVVKKAYQKLCYIHHIMGVKVPDCTACDLVNVLDLDDCIFEQ"

993..2186

CDS

/codon_start=1

/product="Rep 52 protein"

/protein_id="AAC03777.1"

/db_xref="GI:2906020"

/translation="MELVGLVLDKGITSEKOWIQBDQASYISFNAASNSRSQIKAAALD RNTILWGPATTKTNIAEIAHTVPFYGCYVWNTNENFPNDCDQVNIWEEGKMTA KVESAKILGSKVRDQKCKSSAQIDPTPVIIVTSNTNNCAVIDGNTTFEHOQLQ DRMFPELTRLDHDFGKVTQKQVDFFRWAKDHVVEHEFVYKGGAKKRPAPSDA DISEPKRVRESVAQSTSDAESINADRYONKSRHGVNMLFPFCQCRMNQNSNCTFTHGQK DCLECFPVSESQPVSVVKKAYQKLCYIHHIMGVKVPDCTACDLVNVLDLDDCIFEQ"

NAGKIMSLTKTAPDYLVGQOQVEDISSNRIYKILELNGYDPOYAAVSFLGWATKFKG RNTILWGPATTKTNIAEIAHTVPFYGCYVWNTNENFPNDCDQVNIWEEGKMTA KVESAKILGSKVRDQKCKSSAQIDPTPVIIVTSNTNNCAVIDGNTTFEHOQLQ DRMFPELTRLDHDFGKVTQKQVDFFRWAKDHVVEHEFVYKGGAKKRPAPSDA DISEPKRVRESVAQSTSDAESINADRYONKSRHGVNMLFPFCQCRMNQNSN ICFTHGQKDCLECFPVSESQPVSVVKKAYQKLCYIHHIMGVKVPDCTACDLVNVLDL DD CIPEQ"

precursor_RNA

1853..4451

intron

2203..4410

/codon_start=1

/product="major coat protein VP1"

/protein_id="AAC03780.1"

/db_xref="GI:2906023"

/translation="MAADGYLPWLEDTLSEGIQWVKLJPGPPPPKPAERHKDDSRG LVLPGYKVLGPFNGLDKGEVNEADAALAEHDKAYDQLDSGNDPNIYKLNHADAABPOE RKEDTSFGGNLGRAVQAKRVLPLEGLVEEPVKTAPGKRPEVHSPVEDSSGTS KAGQOPARKLNFGQTDGSDVPDPLGQPPPAAPSGLGNTMATGSGAPMADNNEGA DVGNSGNNHCDSTWMDRVTITSTWALPTYNHLYKQISSQSGASNDHHPGVS TPGYFDNFRFCHFSRDMQRLINNNWFRPKLNFKLNIQVKTQNDGTTIAN NLTSVQVFTDSEYOLPYLGSAGHOGCLPPFPADVMVPOYGYLTLANGSOAVGRSSP YCLEYFPQSOLRTGNNFTPSYTPEDVPFPHSSYAHQSOLDRLMPLIDQYLYLRTNT PSQTTQSRLOFSQAGASDIRDSRWLPGCYRQQRVSKTSADNNSEYSWTGTATKY HLNGRDSLVPFGPAMASHKODEEKFPPQGVLIFFGQKSEKINVDIEKWMITDDEIR TTNPAVTEQYGSVSTNLQNRQAATADVTQVLPGMVMDRVDVYLGQPIWAKIPHT DGHFPHPLMGGLGKHPFPQILIKNTPEPANPSTTFAAKFASFITQYSTQVSVETI EMELQKNSKRWNPETQYTSYNKSVNVDFVTDTNGVSEPRPIGTIRLTNL"

2429

variation

/note="compared to sequence of GenBank Accession Number J01901"

/replace="ta"

2614..4410

/transl_except=1

/product="major coat protein VP2"

/protein_id="AAC03778.1"

/db_xref="GI:2906021"

/translation="MAPGKKRPVHSPVPEPSSSGTKAGQOPARKLNFGQTDGADS VPDPQLGQPPAAPSGLGNTMATGSGAPMADNNEGA DVGNSGNNHCDSTWMDRVTITSTWALPTYNHLYKQISSQSGASNDHHPGVS TPGYFDNFRFCHFSRDMQRLINNNWFRPKLNFKLNIQVKTQNDGTTIAN NLTSVQVFTDSEYOLPYLGSAGHOGCLPPFPADVMVPOYGYLTLANGSOAVGRSSP YCLEYFPQSOLRTGNNFTPSYTPEDVPFPHSSYAHQSOLDRLMPLIDQYLYLRTNT PSQTTQSRLOFSQAGASDIRDSRWLPGCYRQQRVSKTSADNNSEYSWTGTATKY HLNGRDSLVPFGPAMASHKODEEKFPPQGVLIFFGQKSEKINVDIEKWMITDDEIR TTNPAVTEQYGSVSTNLQNRQAATADVTQVLPGMVMDRVDVYLGQPIWAKIPHT DGHFPHPLMGGLGKHPFPQILIKNTPEPANPSTTFAAKFASFITQYSTQVSVETI EMELQKNSKRWNPETQYTSYNKSVNVDFVTDTNGVSEPRPIGTIRLTNL"

2429

/note="compared to sequence of GenBank Accession Number J01901"

/replace="ta"

2614..4410

/transl_except=1

/product="major coat protein VP2"

/protein_id="AAC03778.1"

/db_xref="GI:2906021"

/translation="MAPGKKRPVHSPVPEPSSSGTKAGQOPARKLNFGQTDGADS VPDPQLGQPPAAPSGLGNTMATGSGAPMADNNEGA DVGNSGNNHCDSTWMDRVTITSTWALPTYNHLYKQISSQSGASNDHHPGVS TPGYFDNFRFCHFSRDMQRLINNNWFRPKLNFKLNIQVKTQNDGTTIAN NLTSVQVFTDSEYOLPYLGSAGHOGCLPPFPADVMVPOYGYLTLANGSOAVGRSSP YCLEYFPQSOLRTGNNFTPSYTPEDVPFPHSSYAHQSOLDRLMPLIDQYLYLRTNT PSQTTQSRLOFSQAGASDIRDSRWLPGCYRQQRVSKTSADNNSEYSWTGTATKY HLNGRDSLVPFGPAMASHKODEEKFPPQGVLIFFGQKSEKINVDIEKWMITDDEIR TTNPAVTEQYGSVSTNLQNRQAATADVTQVLPGMVMDRVDVYLGQPIWAKIPHT DGHFPHPLMGGLGKHPFPQILIKNTPEPANPSTTFAAKFASFITQYSTQVSVETI EMELQKNSKRWNPETQYTSYNKSVNVDFVTDTNGVSEPRPIGTIRLTNL"

2809..4410

variation

/codon_start=1

/product="major coat protein VP3"

/protein_id="AAC03779.1"

/db_xref="GI:2906022"

/translation="MATGSGAPMADNNEGA DVGNSGNNHCDSTWMDRVTITSTRT WALPTYNHLYKQISSQSGASNDHHPGVS TPGYFDNFRFCHFSRDMQRLINNNW FRPKLNFKLNIQVKTQNDGTTIAN NLTSVQVFTDSEYOLPYLGSAGHOGCL PFPADVMVPOYGYLTLANGSOAVGRSSYCLEYFPQSOLRTGNNFTTFSFDVPP HSSYAHQSOLDRLMPLIDQYLYLRTNTPSQTTQSRLOFSQAGASDIRDSRWL PGPCYRQQRVSKTSADNNSEYSWTGTATKYHLNGRDSLVPFGPAMASHKODEEKFPPQ SVULIFGQKSEKINVDIEKWMITDDEIR TTNPAVTEQYGSVSTNLQNRQAATAD VNTQGVLPGMVMDRVDVYLGQPIWAKIPHTDGHFPHPLMGGLGKHPFPQILIKNTP VPANPSTTFAAKFASFITQYSTQVSVETIEMELQKNSKRWNPETQYTSYNKSVNV DFTVDTNGVSEPRPIGTIRLTNL"

2877

variation

/note="compared to sequence of GenBank Accession Number J01901"

/replace="c"

3759..3765

/note="compared to sequence of GenBank Accession Number J01901"

/replace="g"

3859

/note="compared to sequence of GenBank Accession Number J01901"

```

variation
  J01901" /replace="a"
  3898
  /notes="compared to sequence of GenBank Accession Number
  J01901" /replace="a"
  3900..3902
  /notes="compared to sequence of GenBank Accession Number
  J01901" /replace="gaac"
  4232..4233
  /notes="compared to sequence of GenBank Accession Number
  J01901" /replace="acg"
  4329..4330
  /notes="compared to sequence of GenBank Accession Number
  J01901" /replace="tcg"
  4336
  /notes="compared to sequence of GenBank Accession Number
  J01901" /replace="a"
  4341
  /notes="compared to sequence of GenBank Accession Number
  J01901" /replace="c"
  4347
  /notes="compared to sequence of GenBank Accession Number
  J01901" /replace="t"
  4535..4679
  /notes="inverted terminal repeat"
  /rpt_types=inverted
  misc_feature
  /notes="flop oriented DNA"
  4597..4638
  BASE COUNT 1198 a 1262 c 1255 g 964 t
  ORIGIN

Alignment Scores:
Pred. No.: 1.09e-191 Length: 4679
Score: 2481.50 Matches: 445
Percent Similarity: 90.26% Conservative: 37
Best Local Similarity: 83.33% Mismatches: 51
Query Match: 85.39% Indels: 1
DB: 14 Gaps: 1

US-09-807-802A-17 (1-534) x AF043303 (1-4679)
QY 1 MetAlaSerGlyGlyAlaProMetAlaAspAsnGluGlyAlaAepGlyValGly 20
DB 2809 ATGGCTACAGGAGTGGCGACCAATGGCAGACATTAACGAGGCGCGAGTGGGT 2868
QY 21 AsnAlaSerGlyAenTrpHisCysAspSerThrTrpLeuGlyAepArgValIleThr 40
DB 2869 AATTCTCGGAAATTGGCATTGGATTCCGATTCCACATGATGGCGGACAGATCATCACC 2928
QY 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrIlyvGlnIleSer 60
DB 2929 AGCACCCGAACCTGGGCGCCCTGCCACCTACACACACCACTCTACAAACAAATTTCCAGC 2988
QY 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
DB 2989 CAATCA---GGAGCTCTGAACGACATCACTACTTTGGCTACAGCACCCCTTGGGGGTAT 3045
QY 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
DB 3046 TTTGACTTCAACAGATTCCACTGCCACTTTTCACACACGCTGACTGGCAAGACTCATCAAC 3105
QY 101 AsnAsnTrpGlyPheArgProIlyvAspArgLeuAsnPheIlyvLeuPheAsnIleGlnVal 120
DB 3106 AACAACTGGGGATTCCGACCCCAAGAGACTCAACTTCANGCTCTTTAACTTCAAGTCAA 3165
QY 121 GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 140

```

```

3166 GAGTCACGAGATGACGGTACGACGAGATGTCATTAACCTTACGACGAGGTTCAG 3225
141 ValPheSerAspSerGlyTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
3226 GTGTTTACTGACTCGGAGTACGAGCTCCGCTACGCTCGGCTCGGCGCATCAAGATGC 3285
161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
3286 CTCCCGCGCTTCCGACGAGAGCTTTCATGGTGCCACAGTAGTAGGATACCTTACCCTGAC 3345
181 AenGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlyTyrPheProSerGln 200
3346 AACGGAGTACGACAGTAGGAGCTCTTCATTTACTGCTGGAGTACTTCTCTCTCTCAG 3405
201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
3406 ATGCTGCGTACCGGAAACAACTTTTACCTTACGTACACTTTTGGAGGAGTCTCTTCCAC 3465
221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
3466 AGCAGTACGCTCACAGCCAGAGTCTGGACCGTCTCTATGAATCTCTCTATCGACCATGAC 3525
241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
3526 CTGTATTACTTGAGCAGAACAACTTTCATCTTACGTACACTTTTGGAGGAGTCTCTTCCAC 3585
261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
3586 TTTTCTCAGCGCGGAGCGAGTACATTCGGGACCACTTAGGAATCTGGTCTCTGACCCC 3645
281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
3646 TGTTACCGCCAGCAGCGAGTATCAAGACATCTCGGATACACACACACAGTGAATCTCG 3705
301 TrpThrGlyAlaSerLysTyrAsnLeuAenGlyArgGluSerIleIleAsnProGlyThr 320
3706 TGGACTGGAGTACCAGTACCACCTCAATGGCAGAGACTCTCTGTGTAATCGGCGCCG 3765
321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
3766 GCCATGGCAAGCCACACAGACATGAGAAAGTTTTTTCTCAGAGCGGGTCTTCATC 3825
341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
3826 TTTGGAGCAAGGCTCAGAGAAACAAATGTGGACATTTGAAAGTCTATGATTCAGAC 3885
361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
3886 GAAGAGGAATCAGGACCAACCAATCCCGTGGCTACGGAGCATGTATGTTCTGTATCTACC 3945
381 AsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
3946 AACCTCCAGAGGACGACAGACAGCAGCTACCGCATGTCTACACACACACAGCGTCTT 4005
401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 420
4006 CCAGCATGCTCTGGCAGCAGAGATGTGTACTTTCAGGGGCCCATCTCTGGGCAAGATT 4065
421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
4066 CCACACAGCAGCGGACATTTTCACCCCTCTCCCTCATGGTGGATTCGGACTTAAACAC 4125
441 ProProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460
4126 CTCTCTCCACAGATTCTCATCAAGAACACCCCGGTACTCTCGGATCTCTCGACCACTTC 4185
461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
4186 AGTGGCGGAAAGTTTGTCTTCTCTCATCACAGTACTCCACGAGACAGGTTCAGCGTGGAG 4245
481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrThr 500

```

Db 4246 ATCGAGTGGGAGCTGCGACAGAGAAACAGCAACCGCTCGAATCCGAAATTCAGTACACT 4305
Qy 501 SerAenTyrAlaLeuSerAlaAenValAenPheThrValAenAenGlyLeuTyrThr 520
Db 4306 TCCAACTACACAGCTCTGTTAATGTGACATTACTGTGACACTAATGGCGGTGATATCA 4365
Qy 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
Db 4366 GAGCCTGCCCCATTGGCACCAGATACCTGACTGCTAATCTG 4407

RESULT 10
AF369963
LOCUS AF369963 7327 bp DNA circular SYN 22-JUL-2001
DEFINITION Cloning vector pAAV-RC, complete sequence.
ACCESSION AF369963
VERSION AF369963.1 GI:14994089
KEYWORDS Cloning vector pAAV-RC
SOURCE Cloning vector pAAV-RC
ORGANISM artificial sequences; vectors.
REFERENCE 1 (bases 1 to 7327)
AUTHORS Grafsky, A.J. III.
TITLE PAAV-RC for AAV Helper-Free System
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 7327)
AUTHORS Grafsky, A.J. III.
TITLE Direct Submission
JOURNAL Submitted (16-APR-2001) Technical Services, Stratagene, 11011 N.
Torrey Pines Rd., La Jolla, CA 92037, USA
FEATURES
source
1..7327
/organism="Cloning vector pAAV-RC"
/mol_type="genomic DNA"
/db_xref="taxon:161364"
/lab_host="Escherichia coli"
/note="used for recombinant DNA processing"
131..1996
/note="AAV-2 rep"
/codon_start=1
/transl_table=11
/product="adeno-associated virus 2 rep protein"

CDS

131..1996
/note="AAV-2 rep"
/codon_start=1
/transl_table=11
/product="adeno-associated virus 2 rep protein"
/protein_id="AAK76418.1"
/db_xref="GI:14994090"
/translation="MFGFYEVIVKVPDLDELHPLGIDSDSFVNVAAEKELWELPPDSMDMD
LNLLEQPLTVAELQORLEFTEWRVSKAPEALFVQPEKGSYFHHVLVETTGVS
MVLGFLSQIRLEQLQRYGIEPTLPNFAVTRNGAGGKGVVDRCYIYVYLLPK
TQPELQWATNNEQLSACLNLTERKLVVAQLHIVSTQEQNKENQNPNSDAPVIRS
KTSAYMELGWLVDKGTISEKQIQEDQASTISFNAASNSRSQIKALDNGKIMSL
TKTAPDYLQGPVEDISNRIYKILELNGYDPOYAAVSFVLGATKFGKRNITLFG
PATTKTNIAEALHTVPFGCVNNTNENFPDVCVKWIVWEEGKMTAKVVSASA
ILGSKVRVDQKCSAQIDPTPIVITNTNMCVADGNSTTFEHOQPLQDMFKFEL
TERLDHDFGKVTQKQKDFRWAKDHHVVEHEFVYKKGAKKRPAPSDADISPKRV
RESVAQSTSDAESINADRYQNKSRHVGNLMLFPCRCERMNQNNSNICPTGQK
DCLCFPVSEQSVSVKAYQKJCIHHIMKVPDADTADCLVNVVDLDDCIFIHQ"
2013..4220
/note="cap VP1"
/codon_start=1
/transl_table=11
/product="capsid protein VP1"
/protein_id="AAK76419.1"
/db_xref="GI:14994091"
/translation="MAADQPLDLEDTLSGIRQWKLKPGPPPPKPAERHKDDSRG
LVLPGYKLGPNGLDKGEPVNEADAALSHDKAYDRLDSDGNPPLYKNHADADEFQ
RKQEDTSFGNLRGAVFOAKKRVLEPLGLVEEPVKTAPKKRPVHSPVEPDSSTGTG
KAGQAPARKLNFQGTGADSDVDPQLQGPAPSGLGTMTMATGSPADNNEGA
DGVGNSGNHCDSTWMDRVTITSTRWALPTNNHLYKQISOSGASNDNHYFGYS
TPWGYDFNRHCFSPDRWQLINNPFKRLNFKLFIQVKEVQNDGTITIAN
NLTSTVQVETSEYQLPVLAGSHQCLPPFPADVFMVPOYCYLTLLNNGSOAVGRSSF
YCLEYFPQMLRGNFTFSYTFEDVPFSSVHSSQSLRLNMLPIDOYLVLGRNTY
PSGTTTQSLRQSAGASDIRDOSRNWLPGPCYQQRVSKTSAANNNSYSWTGTKY
HLNGRDSLVPFGPAMASHKDEEKFPPQSGVLIFGKQSQSEKTNVDIEKMLTDEEIR
TTPVATQYSGVSTNLQGNKQAATADVNTQGVLPGMVWDQDVLQGLPIWAKIPIHT
DGHFHPSGVGLGKHPPPQILIKNTVPVNPSTFTTSAKFAFSFITQYSTQGVSVET

CDS

101..1996
/note="AAV-2 rep"
/codon_start=1
/transl_table=11
/product="adeno-associated virus 2 rep protein"
/protein_id="AAK76418.1"
/db_xref="GI:14994090"
/translation="MFGFYEVIVKVPDLDELHPLGIDSDSFVNVAAEKELWELPPDSMDMD
LNLLEQPLTVAELQORLEFTEWRVSKAPEALFVQPEKGSYFHHVLVETTGVS
MVLGFLSQIRLEQLQRYGIEPTLPNFAVTRNGAGGKGVVDRCYIYVYLLPK
TQPELQWATNNEQLSACLNLTERKLVVAQLHIVSTQEQNKENQNPNSDAPVIRS
KTSAYMELGWLVDKGTISEKQIQEDQASTISFNAASNSRSQIKALDNGKIMSL
TKTAPDYLQGPVEDISNRIYKILELNGYDPOYAAVSFVLGATKFGKRNITLFG
PATTKTNIAEALHTVPFGCVNNTNENFPDVCVKWIVWEEGKMTAKVVSASA
ILGSKVRVDQKCSAQIDPTPIVITNTNMCVADGNSTTFEHOQPLQDMFKFEL
TERLDHDFGKVTQKQKDFRWAKDHHVVEHEFVYKKGAKKRPAPSDADISPKRV
RESVAQSTSDAESINADRYQNKSRHVGNLMLFPCRCERMNQNNSNICPTGQK
DCLCFPVSEQSVSVKAYQKJCIHHIMKVPDADTADCLVNVVDLDDCIFIHQ"
2013..4220
/note="cap VP1"
/codon_start=1
/transl_table=11
/product="capsid protein VP1"
/protein_id="AAK76419.1"
/db_xref="GI:14994091"
/translation="MAADQPLDLEDTLSGIRQWKLKPGPPPPKPAERHKDDSRG
LVLPGYKLGPNGLDKGEPVNEADAALSHDKAYDRLDSDGNPPLYKNHADADEFQ
RKQEDTSFGNLRGAVFOAKKRVLEPLGLVEEPVKTAPKKRPVHSPVEPDSSTGTG
KAGQAPARKLNFQGTGADSDVDPQLQGPAPSGLGTMTMATGSPADNNEGA
DGVGNSGNHCDSTWMDRVTITSTRWALPTNNHLYKQISOSGASNDNHYFGYS
TPWGYDFNRHCFSPDRWQLINNPFKRLNFKLFIQVKEVQNDGTITIAN
NLTSTVQVETSEYQLPVLAGSHQCLPPFPADVFMVPOYCYLTLLNNGSOAVGRSSF
YCLEYFPQMLRGNFTFSYTFEDVPFSSVHSSQSLRLNMLPIDOYLVLGRNTY
PSGTTTQSLRQSAGASDIRDOSRNWLPGPCYQQRVSKTSAANNNSYSWTGTKY
HLNGRDSLVPFGPAMASHKDEEKFPPQSGVLIFGKQSQSEKTNVDIEKMLTDEEIR
TTPVATQYSGVSTNLQGNKQAATADVNTQGVLPGMVWDQDVLQGLPIWAKIPIHT
DGHFHPSGVGLGKHPPPQILIKNTVPVNPSTFTTSAKFAFSFITQYSTQGVSVET

EMELQENKRWNPETIQYTSYNKSVNVDFVTDTNGVYSBRPIGTIRYLTRLNL
4705..5161
/note="fl origin"
5292..6152
/note="confers ampicillin resistance"
/codon_start=1
/transl_table=11
/product="beta-lactamase"
/protein_id="AAK76420.1"
/db_xref="GI:14994092"
/translations="MSIQHFRVALIPPPAFCLPVPFHPETLVKVKDAEDQLGARVGY
IEDLNSKILESFRPEERFPMSTFKVLCCGLVDLRIDAGQQLGRIHYSQNDLVE
YSPTKEKLTDMGTVELSAATMTSDNTLANLLLTITIGPKELTAFLEHNGDHTRL
DRPELNEAIPNDERTTTPVAMATLRKLTLCELTLASRQOLIIDWMEADKVGPL
LRSLPAGWFIADKSGAGERSGIITAAALPGDKPSRIIVYITGTSQATMDERRQIA
EIGASLIKHW"
6161..7019
/note="ColE1 origin"
BASE COUNT 1903 a 1887 c 1912 g 1625 t
ORIGIN
Alignment Scores:
Pred. No.: 1.98e-191 Length: 7327
Score: 2481.50 Matches: 445
Percent Similarity: 90.26% Conservative: 37
Best Local Similarity: 83.33% Mismatches: 51
Query Match: 85.39% Indels: 1
DB: 12 Gaps: 1
US-09-807-802a-17 (1-534) x AF369963 (1-7327)
Qy 1 MetAlaSerGlyGlyGlyAlaProMetAlaAspAenGluGlyAlaAspGlyValGly 20
Db 2619 ATGGCTACAGCAGTGGCGCACCAATGCGACACAATTAACGAGGCGCGGAGTGGGT 2678
Qy 21 AsnAlaSerGlyAenTyrHisCysAspSerThrTrpLeuGlyAspArgValIleThr 40
Db 2679 AATTCTCGGAAATTTGGCATTTCGATTCCACATGATGGCGGACAGATCATCACCACC 2738
Qy 41 SerThrArgThrTrpAlaLeuProThrTyrAenAenHisLeuTyrLysGlnIleSerSer 60
Db 2739 AGCACCCGAACCTGGGCGCCCTGCCACCTACAAACACCACTCTTACAAACAATTTCCAGC 2798
Qy 61 AlaSerThrGlyAlaSerAenAspAenHisTyrPheGlyTyrSerThrProTyrGlyTyr 80
Db 2799 CAATCA--GGAGCCTCGAACGACAATCACTACTTTTGGCTACAGCACCCCTTGGGGGTAT 2855
Qy 81 PheAspPheAenArgPheHisCysHisPheSerProArgAspThrGlnArgLeuIleAen 100
Db 2856 TTTGACTTCAACAGATTCCACTGGCCACTTTTCCACCGTGCAGTGGCAAGACTCATCAAC 2915
Qy 101 AsnAenTyrGlyPheArgProLysArgLeuAenPheLysLeuPheAenIleGlnValLys 120
Db 2916 AACAACTGGGATTTCCAGCCCAAGAGACTCAACTTCAAGCTCTTTAACTTCAAGTCAAA 2975
Qy 121 GluValThrThrAenAspGlyValThrThrIleAlaAenAenLeuThrSerThrValGln 140
Db 2976 GAGGTACGCGAGATGACGCTACGCGATTCGCAATAACCTTACCAGCACGGTTCCAG 3035
Qy 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
Db 3036 GTGTTTACTGACTCGGAGTACCAGCTCCCGTACCTCCCTCGGCTCGGCGCATCAAGGATGC 3095
Qy 161 LeuProProPheProAlaAenValPheMetIleProGlnTyrGlyTyrLeuThrLeuAen 180
Db 3096 CTCCCGCGTTCGCCAGAGACGCTTCTCATGTGTCACAGATGATGATACCTCACCTGAAAC 3155
Qy 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
Db 3156 AACGGAGTACGAGCAGTAGGACGCTCTTCAATTTTACTGCTGGAGTAGTACTTTCTTCTCAG 3215
Qy 201 MetLeuArgThrGlyAenAenPheThrPheSerTyrThrPheGluGluValProPheHis 220

Db 3554 CTGTATTACTTGAGCAGAACAAACACTCCAAAGTGGAAACACCACCGCAGTCAAGGCTTCAG 3613
 Qy 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
 Db 3614 TTTTCTCAGCCGCGAGTGACATTCGGACCACTAGGACCTTGGCTTCTCGACCC 3673
 Qy 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
 Db 3674 TGTATCCGCCAGCAGCGAGTATCAAGACATCTCGGATTAACAACAACAGTGAATCTCG 3733
 Qy 301 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr 320
 Db 3734 TGGACTGGAGCTACCAAGTACCCTCAATGGCAGAGACTCTGTGTGAATCCGGCCCG 3793
 Qy 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
 Db 3794 GCCATGGCAAGCCACAGGACGATGAAGAAAGTTTTTCTCAGAGCGGGTCTCATC 3853
 Qy 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
 Db 3854 TTTGGGAAGCAAGGCTCAGAGAAACAAATGTGGACATTCMAAAGGTCTATGATTACAGAC 3913
 Qy 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
 Db 3914 GAAGAGGAATCAGGCAACCAATCCCGTGGCTACGGAGCGATGTTCTGTATATCC 3973
 Qy 381 AsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
 Db 3974 AACCTCCAGAGGCCACAGACAGCAGCTACCGCAGATGTCAACACACAGCGTCTT 4033
 Qy 401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 420
 Db 4034 CCAGGATGCTCTGGCAGGACAGAGATGTGTACTTTCAGGGGCCCATCTGGGCAAGATT 4093
 Qy 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
 Db 4094 CCACACGAGGACGACATTTTCAACCTCTCCCTCATGGTGGATTTCGAGCTTAAACAC 4153
 Qy 441 ProProGlnIleLeuIleLysAsnThrProValProAlaAsnProAlaGluPhe 460
 Db 4154 CCTCTCTCCAGATTCTCATCAGAACACCCCGTACCTGGCAATCTTCGACCCTTC 4213
 Qy 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
 Db 4214 AGTGGCGCAAGTTTGTCTTCTCATCACACAGTACTCCACGGGACAGGTCAGCGTGGAG 4273
 Qy 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrThr 500
 Db 4274 ATCGAGTGGAGCTGCAGAAAGGAAACAGCAACGCTGGAATCCCGAAATTCAGTACACT 4333
 Qy 501 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 520
 Db 4334 TCCAACTACAAACAGTCTGTATATGTGGACTTTTACTGTGGACACTAATGGCGTGATTCA 4393
 Qy 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
 Db 4394 GAGCCTCGCCCAATGGCCACAGATACCTGACTCGTAATCTG 4435

RESULT 12

AX205073 LOCUS 7557 bp DNA linear PAT 30-AUG-2001
 DEFINITION Sequence 3 from Patent WO0155361.
 ACCESSION AX205073

VERSION AX205073.1 GI:15394314

SOURCE unidentified

ORGANISM unidentified

REFERENCE 1

AUTHORS Hardy, S. F.

TITLE Recombinant aav packaging systems

JOURNAL Patent: WO 0155361-A 3 02-AUG-2001;

Chiron Corporation (US)

FEATURES

source

Location/Qualifiers

1..7557
 /organism="unidentified"
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"
 /note="recombinant DNA"

BASE COUNT 1970 a 1940 c 1953 g 1694 t

ORIGIN

Alignment Scores:

Pred. No.: 2,076-191 Length: 7557
 Score: 2481.50 Matches: 445
 Percent Similarity: 59.26% Conservative: 37
 Best Local Similarity: 83.33% Mismatches: 51
 Query Match: 85.39% Indels: 1
 DB: 6 Gaps: 1

US-09-807-802A-17 (1-534) x AX205073 (1-7557)

Qy 1 MetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGly 20
 Db 2837 ATGGCTACAGGCAGTGGCGCCACCAATGCGACACAATAACGAGGGCGCGAGTGGGT 2896
 Qy 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThr 40
 Db 2897 AATTCCTCGGAAAATTGGCATTTCCGATTTCCACATGGATGGGCGCAGAGTCTATCCACC 2956
 Qy 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSer 60
 Db 2957 AGCACCAGAACCTGGGCGCTGCCACCTACAAACACCACTCTCAACAACAAATTTCCAGC 3016
 Qy 61 AlaSerThrGlyAlaSerAsnAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
 Db 3017 CAATCA---GGAGCCTCGAACGACAATCACTACTTTTGGCTACAGCACCCCTTGGGGGTAT 3073
 Qy 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
 Db 3074 TTTGACTTCAACAGATTCCACTGCCACTTTTCCACCGTGACTGGCAAGACTCATCAAC 3133
 Qy 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 120
 Db 3134 AACAACTGGGGAATCCGACCCACAGAGACTCAACTTCAAGCTCTTTAACAATTCAGTCAAA 3193
 Qy 121 GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 140
 Db 3194 GAGGTCAACGAGAATGACGGTACGACGACGATTTGCCAATAACCTTACAGCACGGTTCAG 3253
 Qy 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
 Db 3254 GTGTTTACTGACTCGGAGTACGAGCTCCGCTACGCTCGGCTCGGCGCATCAAGGATGC 3313
 Qy 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
 Db 3314 CTCGCCCGGTTCCAGCAGACAGCTTTCATGTGGCCACAGTATGGATATCTCACTCCCTGAA 3373
 Qy 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
 Db 3374 AACGGGAGTCAGGCAGTAGGACGCTCTCATTTTACTGCTGGAGTACTTTCTCTCTCAG 3433
 Qy 201 MetLeuArgThrGlyLysAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
 Db 3434 ATGCTGCGTACCGGAACAACATTTTACCTTACGTACACTTTTGGAGGAGCGTTCCTTCCAC 3493
 Qy 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
 Db 3494 AGCAGCTACGCTCACAGCAGAGTCTGGACCGCTCTCATGATCTCTCATCCACCCAGTAC 3553
 Qy 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
 Db 3554 CTGTATTACTTGGAGCAGAAACAACTCCAAGTGGAAACCCACCGCAGTCAAGGCTTCAG 3613
 Qy 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280

Db 3614 TTTTCTCAGCGCGGAGGTGACATTCGGGACCATCTAGGAACTGGCTTCTCGACCC 3673
Qy 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
Db 3674 TGTATCCCGCAGCAGCGAGTATCAAGACATCTCGCGATAACAACAACAGTGAATCTCG 3733
Qy 301 TTPThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleLeuAsnProGlyThr 320
Db 3734 TGGACTGGAGCTACCAAGTACCACTCAATGGCAGAGACTCTCTGGTGAATCCGGGCCCG 3793
Qy 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
Db 3794 GCATGGCAGCCACACAGGACGATGAAGAAAGTTTCTCAGAGCGGGTTCATC 3853
Qy 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
Db 3854 TTTGGGAAGCAAGCTCAGAGMAAACAATGTGGACATTTGAAAGGTTCATGATTACAGAC 3913
Qy 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
Db 3914 GAAGAGAAATCAGGACAACTCCCGTGGCTACGGAGCAGTATGGTTCTGTATCTACC 3973
Qy 381 AsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
Db 3974 AACCTCAGAGAGCAACACAGCAGCTACCGCAGATGTCAACACACAAGCGTCTT 4033
Qy 401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 420
Db 4034 CAGGCATGCTTGGCAGGACAGAGATGTGTACCTTCAGGGGCCCATCTCGGCAAGATT 4093
Qy 421 ProHisThrAspGlyHisPheHisProSerProleuMetGlyGlyPheGlyLeuLysAsn 440
Db 4094 CCACACAGGACGAGACATTTTCACCCCTCTCCCTCACTGGGTGGATTTCGGACTTAAAC 4153
Qy 441 ProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460
Db 4154 CCTCTCCACAGATTCTCATCAAGAACACCCCGGTACTCTCGAATCCTTCGACCACTTC 4213
Qy 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
Db 4214 AGTCGGCAAGATTGTCTTCTCATCACACAGTACTCCACGGGACAGGTTCAGGTGGAG 4273
Qy 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrThr 500
Db 4274 ATCAGTGGGAGCTGCGAGAGAGAAACAGCAACCGCTGGAATCCGAAATTCAGTACACT 4333
Qy 501 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 520
Db 4334 TCCAACCTACAACAGTCTGTTAATGTGGACTTTTACTGTGGACACTAATGGCGTATTCA 4393
Qy 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
Db 4394 GAGCTCGCCCATGGCCACCATACCTGACTCGTAATCTG 4435

RESULT 13

LOCUS AR222044 8698 bp DNA linear PAT 26-SEP-2002
DEFINITION Sequence 2 from patent US 6429001.
ACCESSION AR222044
VERSION AR222044.1 GI:23329400

KEYWORDS

Unknown.

SOURCE

Unknown.

REFERENCE

1 (bases 1 to 8698)

Hardy, S.F.

Recombinant AAV packaging systems

Patent: US 6429001-A 2 06-AUG-2002;

Location/Qualifiers

1. 8698

/organism="unknown"

BASE COUNT 2117 a 2368 c 2306 g 1907 t

ORIGIN

Alignment Scores:
Pred. No.: 2,498-191 Length: 8698
Score: 2481.50 Matches: 445
Percent Similarity: 90.26% Conservative: 37
Best Local Similarity: 83.33% Mismatches: 51
Query Match: 85.33% Indels: 1
DB: 6 Gaps: 1
US-09-807-802A-17 (1-534) x AR222044 (1-8698)
Qy 1 MetAlaSerGlyGlyGlyValAlaProMetAlaAspAsnAsnGluGlyValAlaAspGlyValGly 20
Db 2809 ATGGCTACAGGAGTGGCGCCACCATGCGACAGACATACAGAGCGCGCGAGTGGGT 2868
Qy 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThr 40
Db 2869 AATTCTCGGAAATTTGGCATTTCCGATTCACATGGATGGCGCAGAGTTCATCCACC 2928
Qy 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSer 60
Db 2929 AGCACCCGAACTGGGCCCTGCCACCTACAAACACCCTCTACAAACAAATTTCCACG 2988
Qy 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
Db 2989 CAATCA--GGAGCTTCGACGACCATCACTACTTTGGCTACAGCACCCCTTGGGGGTAT 3045
Qy 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
Db 3046 TTTGACTTCAACAGATTCACCTGCCACTTTTCCACCGTGACTGGCAAGACTCATCAAC 3105
Qy 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 120
Db 3106 AACAACTGGGGATTCCGACCCAGAGACTCACTTCAAGCTCTTTACACTTCAGTCAA 3165
Qy 121 GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 140
Db 3166 GAGGTACCCAGCATGACGGTACGACGACGATGGCCATTAACCTTACCAGCAGCGTTCAG 3225
Qy 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
Db 3226 GTGTTTACTGACTCGGAGTACAGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCG 3285
Qy 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
Db 3286 CTCCGCCCTTCCAGCAGACGCTTTCATGGTCCACAGTATGGATACCTCCACCTGAC 3345
Qy 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
Db 3346 AACGGGAGTCAGCAGTACGACGCTCTTCACTTACTGCTGGAGTACTTCTCTCTCTCAG 3405
Qy 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
Db 3406 ATGCTCGGTACCGGAACAACCTTTACCTTACCTACACTTTTGGAGGACGTTCTCTTCCAC 3465
Qy 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
Db 3466 AGCAGCTACCTCAGCAGCAGGAGTCTGGACCGCTCATGAACTCTCTCATCGACCATGAC 3525
Qy 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
Db 3526 CTGTATTACTTGAGCAGAAACAACACTCCCAAGTGGAAACCAACCAACGAGTCAAGGCTTCAG 3585
Qy 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
Db 3586 TTTTCTCCGCGCGGAGCGAGTGACATTCGGGACCATGCTAGGAACTGGCTTCTCTGACCC 3645
Qy 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
Db 3646 TGTATCCGCGCAGCGAGTATCAAGAGATCTGCGGATAACACACACAGTGAATCTCG 3705
Qy 301 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleLeuAsnProGlyThr 320

```
Db 3706 TGGACTGGAGCTACCAAGTACACCTCAATGGCAGAGACTCTCTGGTGAATCCCGGCCG 3765
Qy 321 AlaMetAlaSerHisLysAspGluAspLysPhePheProMetSerGlyValMetile 340
Db 3766 GCCATGGCAAGCCACAGGACGATGAAGAAAGTTTTTCTTCAGAGGGGGTCTCATC 3825
Qy 341 PheGlyLysGluSerAlaGlyAlaSerAenThrAlaLeuAspAenValMetileThrAsp 360
Db 3826 TTTTGGGAAGCAAGGCTCAGAGAAAACAATGTGGACATTTGAAGAGTTCATGATTACAGAC 3885
Qy 361 GluGluGluIleLysAlaThrAenProValAlaThrGluArgPheGlyThrValAlaVal 380
Db 3886 GAAGAGGAATCAGACAAACCATCCCTGGCTACGGACGATATGGTTCGTATCTACC 3945
Qy 381 AenPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
Db 3946 AACCTCCAGAGAGGCAACAGACAGCAGCTACCGCAGATGTCAACACACAAGGCGTCTT 4005
Qy 401 ProGlyMetValTTPGlnAspArgAspValTyrLeuGlnGlyProIleThrAlaLysile 420
Db 4006 CCAGGCATGGTCTGGCAGACAGAGATGTGTACCTTTCAGGGGCCCATCTGGGCAAGATT 4065
Qy 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAen 440
Db 4066 CCACACAGGACGGACATTTTCCACCCCTCTCCCTCATGGGTGGATTCCGACTTAAACAC 4125
Qy 441 ProProGlnIleLeuIleLysAenThrProValProAlaAenProProAlaGluPhe 460
Db 4126 CCTCTCCACAGATTTCTCATCAAGAACACCCCGGTACCTGGGAATCCTTCGACCACTTC 4185
Qy 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
Db 4186 AGTGGCGCAAGTTTGGCTTCTTCATCAACAGTACTCCAGGGACAGGTTCAGCGTGGAG 4245
Qy 481 IleGluTTPGluLeuGlnLysGluAenSerLysArgTTPAenProGluValGlnTyrThr 500
Db 4246 ATCGAGTGGAGCTCAGAGAGAAACAGCAACCTGGATTCGCCAAATTCAGTACACT 4305
Qy 501 SerAenTyrAlaLysSerAlaAenValAspPheThrValAspAenAenGlyLeuTyrThr 520
Db 4306 TCCAACTACAAACAGTCTGTTAATGTGGACTTTTACTGTGGACACTAATGGCGTGATTCA 4365
Qy 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
Db 4366 GAGCCTCGCCCCATTGGCACCAGATACCTGACTCGTAATCTG 4407
```

```
RESULT 14
AX205072
LOCUS AX205072 8698 bp DNA linear PAT 30-AUG-2001
DEFINITION Sequence 2 from Patent WO0155361.
ACCESSION AX205072
VERSION AX205072.1 GI:15394313
KEYWORDS
SOURCE unidentified
ORGANISM unidentified
unclassified.
```

```
REFERENCE
1. Hardy, S.F.
AUTHORS Recombinant aav packaging systems
TITLE Patent: WO 0155361-A 2 02-AUG-2001;
JOURNAL Chiron Corporation (US)
FEATURES
location/Qualifiers
source
1. 8698
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/note="recombinant DNA"
```

```
BASE COUNT 2117 a 2368 c 2306 g 1907 t
ORIGIN
```

```
Alignment Scores:
Pred. No.: 2,49e-191 Length: 8698
Score: 2481.50 Matches: 445
```

```
Percent Similarity: 90.26% Conservative: 37
Best Local Similarity: 83.33% Matches: 51
Query Match: 85.39% Indels: 1
DB: 6 Gaps: 1
US-09-807-802a-17' (1-534) x AX205072 (1-8698)
Qy 1 MetAlaSerGlyGlyGlyAlaProMetAlaAspAenGluGlyAlaAspGlyValGly 20
Db 2809 ATGGCTACAGCAGTCGCGCACCAATTCGACAGACAATAACGAGGGCGCGGAGTGGGT 2868
Qy 21 AenAlaSerGlyAenTTPHisCysAspSerThrTTPLeuGlyAspArgValIleThrThr 40
Db 2869 AATTCTCTCGGAAATTCGCAATTCGATTCACATGGATGGCGACAGAGTTCATCCACC 2928
Qy 41 SerThrArgThrTTPAlaLeuProThrTyrAenAenHisLeuTyrLysGlnIleSerSer 60
Db 2929 AGCACCCGACCTGGGCGCTTCCACCTACACCAACCACTCTTACAAACAATTTCCAGC 2988
Qy 61 AlaSerThrGlyAlaSerAenAenHisTyrPheGlyTyrSerThrProTTPGlyTyr 80
Db 2989 CAATCA--GGAGCCTCGAACGACCAATCACTACTTTGGCTACAGCACCCCTTGGGGTAT 3045
Qy 81 PheAspPheAenArgPheHisCysHisPheSerProArgAspTTPGlnArgLeuIleAen 100
Db 3046 TTTGACTTCAACAGATTCCTCACTGCCACTTTTACCACGTGACTGGCAAGACTCATCAAC 3105
Qy 101 AenAenTTPGlyPheArgProLysArgLeuAenPheLysLeuPheAenIleGlnValLys 120
Db 3106 AACAACTGGGGATTCGACCCACAGAGACTCACTCAAGCTCTTTAACAATTCAGTCAAA 3165
Qy 121 GluValThrThrAenAspGlyValThrIleAlaAenAenLeuThrSerThrValGln 140
Db 3166 GAGGTCAACGAGATGACGGTACGACGAGATTGCCAATAACCTTACCAGCACGGTTCAG 3225
Qy 141 ValPheSerAspSerGlyTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
Db 3226 GTGTTTACTGACTCGGAGTACCGACTCCGCTACCTCGGCTCGGCGCATCAAGGATGC 3285
Qy 161 LeuProProPheProAlaAspValPheMetileProGlnTyrGlyTyrLeuThrLeuAen 180
Db 3286 CTCCCGCGTTCACGACGAGCTTTCATGGTGCCACAGTATGGATACCTCACCTTGAC 3345
Qy 181 AenGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlyTyrPheProSerGln 200
Db 3346 AACGGAGTCAAGCAGTAGGACGCTCTTCACTTTTACTGCTGGAGTACTTTCTCTCTCAG 3405
Qy 201 MetLeuArgThrGlyAenAenPheThrPheSerTyrThrPheGluGluValProPheHis 220
Db 3406 ATGCTGGTACCGGAAACAACCTTTTACCTTACGCTACACTTTTGGAGGAGTTCCTTTCCAC 3465
Qy 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAenProLeuIleAspGlnTyr 240
Db 3466 AGCAGCTACGCTCACGACGAGAGTCTGAGCGCTCTCATGAATCTCTCATCGACCACTAC 3525
Qy 241 LeuTyrTyrLeuAenArgThrGlnAenGlnSerGlySerAlaGlnAenLysAspLeuLeu 260
Db 3526 CTGTATTACTTGAGCAGAACAAACACTCCAAGTGGAAACCCAGCCAGTCAAGGCTTCAG 3585
Qy 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAenTTPLeuProGlyPro 280
Db 3586 TTTTCTCAGGCGCGGAGCGAGTACATTTCGGGACCACTTAGGAACCTGGCTTCTTGGACCC 3645
Qy 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAenAenAenSerAenPheThr 300
Db 3646 TGTATACCGCCAGCAGCGAGTATCAAAGACATCTCGGATTAACAACAACAGTGAATACTCG 3705
Qy 301 TTPThrGlyAlaSerLysTyrAenLeuAenGlyArgGluSerIleIleAenProGlyThr 320
Db 3706 TGGACTGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTGTGTAATCCGGGCGCG 3765
Qy 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetile 340
```

```
Db 3766 GCATGGCAAGCCACAGGACGATGAAGAAAGTTTTTCTCAGAGCGGGTCTCATC 3825
Qy 341 PheGlyLeuSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
Db 3826 TTTGGGAAGCAAGGCTCAGAGAAACAAATGTGGACATTGAAAGGTCATGATTACAGAC 3885
Qy 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
Db 3886 GAAGAGGAATCCAGCAACCAATCCCGTGGCTACGGAGCAGATGTTCTGTATCTACC 3945
Qy 381 AsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
Db 3946 AACTCCAGAGGCAACACACAGCAGCTACCGCAGATGTCAACACACAGCGGTTCTT 4005
Qy 401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleThrAlaLysIle 420
Db 4006 CCAGGATGCTCTGGCAGGACAGAGATGTGTACTCTTCAGGGGCCCATCTGGCCAAAGATT 4065
Qy 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
Db 4066 CCACACAGGACGACATTTTACCCCTCTCCCTCATGGGTGGATTCGACTTAAACAC 4125
Qy 441 ProProGlnIleLeuLysAsnThrProValProAlaAsnProProAlaGluPhe 460
Db 4126 CCTCTCCACAGATTCTCATCAAGACACCCCGTACCTGGGAATCCTTCGACCACTTC 4185
Qy 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
Db 4186 AGTGGCGCAAGTTTGTCTCTTCATCACACAGTACTCCACGGGACAGGTGCGTGGAG 4245
Qy 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrThr 500
Db 4246 ATCAGTGGGAGCTGCAGAGGAAGAACAGCAACGCTGGGAATCCCGAATTCAGTACACT 4305
Qy 501 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 520
Db 4306 TCCAACTACAAAGTCTGTTAATGTGGACTTTTACTGTGGACACTAATGGCGGTATTCA 4365
Qy 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
Db 4366 GAGCCTCGCCCATTTGGCACCAGATACCTGACTCGTAATCTG 4407

RESULT 15
LOCUS AY243021 2208 bp DNA linear VRL 14-MAY-2003
DEFINITION Non-human primate Adeno-associated virus isolate AAVch.5 capsid
protein (VP1) gene, complete cds.
ACCESSION AY243021
VERSION AY243021.1 GI:29650537
KEYWORDS Non-human primate Adeno-associated virus
SOURCE Non-human primate Adeno-associated virus
ORGANISM Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
REFERENCE 1 (bases 1 to 2208)
AUTHORS Gao, G., Alvira, M.R., Somanathan, S., Lu, Y., Vandenbergh, L.H.,
Rux, J.J., Calcedo, R., Sammiguel, J., Abbas, Z. and Wilson, J.M.
TITLE Adeno-associated viruses undergo substantial evolution in primates
during natural infections
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (10), 6081-6086 (2003)
PUBMED 12716974
REFERENCE 2 (bases 1 to 2208)
AUTHORS Gao, G., Alvira, M.R., Somanathan, S., Lu, Y., Vandenbergh, L.H.,
Rux, J.J., Calcedo, R., Sammiguel, J., Abbas, Z. and Wilson, J.M.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-2003) Department of Medicine, University of
Pennsylvania School of Medicine and Wistar Institute, 204 Wistar
Institute, 3601 Spruce Street, Philadelphia, PA 19104, USA
FEATURES
source
1..2208
/organism="Non-human primate Adeno-associated virus"
/proviral
/mol type="genomic DNA"
/isolate="AAVch.5"
```

```
gene /db_xref="taxon:226502"
1..2208
CDS /gene="VP1"
1..2208
/gene="VP1"
/codon_start=1
/product="capsid protein"
/protein_id="AAO8207.1"
/db_xref="GI:29650538"
/translation="MAADGYLPDWMLEDTLSGIRQWKLKPGPPPKPNQQRDDSRG
LVLPGLYKLPFNGLDKGEVNEADAALAEHDKAYDHLQKQDNDPILKYNHADAQFQE
RLOEDTSFGNLRGAVFQAKRVLPLEGLVEAVKTAPEKRPISQSPAPDSSGIG
KSGQPAKRLNFGQGTGTEGVPAPQPIGEPPAAPSGVSNMAGSGAPADNNEGA
DVGNSGNWCHDSTWMDRVLTTSTRWALPTNNHLYKFISSSGATNDNHGYEAS
TPWGYPDRFCHPESPRDWRLNNNGWPKLNPFLKNOVKEVTONDQTTTIAN
NLTSVQVTDSEYQLPVLGSAHQCLPPPADVFEMIPQVGLTLNNGSOAVGSSP
TCLYFFPSQMLRTGNFTSYFEDVPFHSYAHQSQSLRLNPLDILQYLYLSTQG
YSGTQQSRLQFSQAPSSMAQAKWLPSPSYRQORSMKTANDNNSEFAWTAATKY
YLNGRNSLVNPPPMASHKDDEKYFPMHGNLIFGKQGTGTNNVDIEVLITDEEIR
TTNPVATQYGOVATNHQSNTASYSDVDSQGLPGMWQDRDVLQOPTWAKTPT
DGHFHPSLMGFGGLKHPPLQILLKNTVPANPATFTTGPKEASFTIYSGOVSEI
EWELOKNSKRWNPFIQTSYNSKSNVVEFTVDANGVYSRPIGTRILTRNL"
BASE COUNT 617 a 638 c 518 g 435 t
ORIGIN
Alignment Scores:
Pred. No.: 3,6e-190 Length: 2208
Score: 2457.50 Matches: 442
Percent Similarity: 89.70% Conservative: 37
Best Local Similarity: 82.77% Mismatches: 54
Query Match: 84.57% Indels: 1
DB: 14 Gaps: 1
US-09-807-802A-17 (1-534) x AY243021 (1-2208)
Qy 1 MetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGly 20
Db 607 ATGGCTTCAGGCGGTGGGCGACCAATGGCAGACAATAACGAAGGCGCGCAGGATGGGT 666
Qy 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 40
Db 667 AATTCCTCGGAAATTTGGCATTGCCATTCACATGGATGGGCGCAGAGTTATCACCACC 726
Qy 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSer 60
Db 727 ACACACAGAACCTGGCGCTCCCGCTCCACCTACATATATACCTCTACAGCAATCTCCAGC 786
Qy 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
Db 787 GAATCG--GGAGCCACCAACGACCACTACTTTCGGCTACAGCACCCCTCGGGGGTAT 843
Qy 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
Db 844 TTTGACTTTTAAACAGATTCCACTGTCTCTCACCACGTCGACTGGCAGGACTCATCAAC 903
Qy 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 120
Db 904 AACACTGGGGATTAGACCCCAAGAACTCAATTTCAAGCTCTTCAACATCCAAAGTCAAG 963
Qy 121 GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 140
Db 964 GAGGTCAACGAGAAATGATGGAACACACGACCATCGCCAAATAACCTTACCAGCAGCGTGCAG 1023
Qy 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
Db 1024 GTCTTCACAGACTCTGAGTACAGCTGCCCTACGTCTCGGTTCGGCTCACCAGGCTGC 1083
Qy 161 LeuProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
Db 1084 CTTCCGCGCTTCCAGCAGACGCTTTCATGATTCTCTCAGTACGGCTACTTGTCTCTGAAC 1143
Qy 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
```

Db 1144 AATGCACGCCAAGCGGTAGGACGTTCTTCATCTCTACTCTAGAGTATTTTCCCTCTCAG 1203
Qy 201 MetLeuArgThrGlyAsnAsnPhetheThrPheSerTyrThrPheGluGluValProPheHis 220
Db 1204 ATGCTGAGGACGGGAACAACATTCACCTTCAGCTACACTTTTGAAGACGTCCTTTCCAC 1263
Qy 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuLeuAspGlnTyr 240
Db 1264 AGCAGCTACGGGCACAGCCAGAGTCTGGATCGGCTGATGAATCTCTCTCATTTGACCACTAC 1323
Qy 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
Db 1324 CTGTATTACCTGAGCAAACTCAGGGTACAGTGGACACACGACGACAACTCGAGACTGCAG 1383
Qy 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
Db 1384 TTCACCCAAGCTGGGCTAGTCCATGCTCAGCAGCGCCAAAACCTGGCTACCGGACCC 1443
Qy 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPhetheThr 300
Db 1444 AGCTACCGACAGCAGCAATGTCTAAGACGGCTAATGACACCAACACACAGTGAATTTGCT 1503
Qy 301 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr 320
Db 1504 TGGACTGACGCCACCACAATATTACTCTGAATGGAATAATCTCTGGTCAATCCCGGGCCC 1563
Qy 321 AlaMetAlaSerHisLysAspGluAspLysPhePheProMetSerGlyValMetIle 340
Db 1564 CCAATGGCCAGTCACACAGGACGATCAGGAAAGTATTTCCCATGCACGGAATCTCATC 1623
Qy 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
Db 1624 TTTGGAAAACAAGGCACAGCACTCAATGTGGACATTTGAATCAGTGTCTTATTACAGAC 1683
Qy 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
Db 1684 GAAGAAGAAATCAGAACACTATCTGTGGTACAGAACATACGGACAGGTGGCCACC 1743
Qy 381 AsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
Db 1744 AACCATCAGAGTCAGACACACACACCTTCCTATGGAAGTGTGGACAGCCAGGGAATCTTA 1803
Qy 401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaIleVal 420
Db 1804 CCTGGAATGGTGTGGCAGACCGCGATGTCTATCTTCAAGGTCCCATTTGGGCCCAAACT 1863
Qy 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
Db 1864 CCTCACGCGCGGACATTTTCATCTCTCCGCTCATGGAGGCTTTGGACTGAACAC 1923
Qy 441 ProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460
Db 1924 CTTCTCTCCCGAGATCCTGATCAAAAACACACCTGTGCCGCGAATCCCGCGACCACTTTC 1983
Qy 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
Db 1984 ACTCTGGAAAGTTTGTCTCGTTCATTACCCAGTATTCACCGGACAGGTGAGCGTGGA 2043
Qy 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrThr 500
Db 2044 ATAGAGTGGAGCTCGAGAGAGAAAACAGCAACCTGTGACCCAGAAATTCAGTACACC 2103
Qy 501 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 520
Db 2104 TCCAACACTACAACAGTCGGTGAATGTGGATTTTACCGTGGACGCAACCGGTGTTTATTCT 2163
Qy 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
Db 2164 GAACCCCGCCCTATTGGCAGCTCGTTACCTTACCCGGAACCTTG 2205

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 14, 2004, 15:18:00 ; Search time 366.286 Seconds
(without alignments)
3935.452 Million cell updates/sec

Title: US-09-807-802A-17

Perfect score: 2906

Sequence: 1 MASGGAPADNNEGADGVG.....NNGLYTEPRIGTRYLTRPL 534

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q/cn2_1/USFTO_spool/US09807802/runat_11022004.175606.15919/app_query.fasta_1.2389
-DB=N Geneseq 15Jun03 -QFT=fastp -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALFMT=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09807802 @CN 1.1 806 @runat_11022004.175606.15919 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq 19Jun03: *
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT: *
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT: *
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT: *
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT: *
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT: *
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT: *
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT: *
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT: *
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT: *
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT: *
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT: *
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT: *
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT: *
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT: *
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT: *
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT: *
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT: *
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT: *
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT: *
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT: *
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT: *
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT: *
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT: *
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT: *
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|------------------------------|
| 1 | 2906 | 100.0 | 1605 | 21 | AAD00779 Adeno-associated v |
| 2 | 2906 | 100.0 | 1800 | 21 | AAD00778 Adeno-associated v |
| 3 | 2906 | 100.0 | 2211 | 21 | AAD00777 Adeno-associated v |
| 4 | 2906 | 100.0 | 4718 | 21 | AAD00772 Adeno-associated v |
| 5 | 2884 | 99.2 | 4683 | 22 | AAF23749 AAV6 DNA sequence. |
| 6 | 2530 | 87.1 | 4722 | 22 | AAF23748 AAV3B DNA sequence. |
| 7 | 2518 | 86.6 | 8178 | 21 | AAD00834 ChimERIC adeno-ass |
| 8 | 2481.5 | 85.4 | 4072 | 22 | AAD00832 Adenovirus helper |
| 9 | 2481.5 | 85.4 | 4679 | 22 | AAD00831 Adeno-associated v |
| 10 | 2481.5 | 85.4 | 4679 | 22 | AAF23750 AAV2 DNA sequence. |
| 11 | 2481.5 | 85.4 | 4679 | 24 | ABK89694 Adeno-associated v |
| 12 | 2481.5 | 85.4 | 4679 | 25 | ABV76133 Adeno-associated v |
| 13 | 2481.5 | 85.4 | 7557 | 22 | AAH26326 Adeno-associated v |
| 14 | 2481.5 | 85.4 | 8698 | 22 | AAH26325 Wild-type adeno-as |
| 15 | 2439.5 | 83.9 | 4680 | 17 | AAT09008 Wild-type adeno-as |
| 16 | 2439.5 | 83.9 | 4680 | 25 | ABX14497 Wild-type adeno-as |
| 17 | 2423.5 | 83.4 | 4675 | 22 | AAH41481 Adeno-associated v |
| 18 | 2423.5 | 83.4 | 4675 | 22 | AAF89931 Nucleotide sequenc |
| 19 | 2423.5 | 83.4 | 4675 | 24 | ABS69879 Human adeno-associ |
| 20 | 2423.5 | 83.4 | 4675 | 24 | ABS69880 Human adeno-associ |
| 21 | 2423.5 | 83.4 | 4675 | 24 | ABA02989 Adeno-associated v |
| 22 | 2258.5 | 77.7 | 8151 | 21 | AAD00832 ChimERIC adeno-ass |
| 23 | 1690.5 | 58.2 | 1617 | 19 | AAV21658 AAV4 VP3 coat prot |
| 24 | 1690.5 | 58.2 | 1617 | 25 | ABX15817 Adeno-associated v |
| 25 | 1690.5 | 58.2 | 1800 | 19 | AAV21657 AAV4 VP2 coat prot |
| 26 | 1690.5 | 58.2 | 1800 | 25 | ABX15816 Adeno-associated v |
| 27 | 1690.5 | 58.2 | 2208 | 19 | AAV21650 AAV4 VP1 capsid pr |
| 28 | 1690.5 | 58.2 | 2208 | 25 | ABX15809 Adeno-associated v |
| 29 | 1690.5 | 58.2 | 4767 | 19 | AAV21648 AAV4 genome. Aden |
| 30 | 1690.5 | 58.2 | 4767 | 25 | ABX15807 Adeno-associated v |
| 31 | 1690.5 | 58.2 | 2487 | 18 | AAT69854 Duck parvovirus ca |
| 32 | 1680.5 | 57.8 | 7215 | 21 | AAD00831 Hybrid adeno-associ |
| 33 | 1676 | 57.7 | 4570 | 16 | AAT01190 Barbary duck parvo |
| 34 | 1665 | 57.3 | 2264 | 21 | AZA49213 Adeno associated v |
| 35 | 1665 | 57.3 | 2264 | 21 | AZA49214 Adeno associated v |
| 36 | 1665 | 57.3 | 2307 | 21 | AAZ49212 Adeno associated v |
| 37 | 1665 | 57.3 | 4652 | 21 | AAZ49209 Adeno associated v |
| 38 | 1665 | 57.3 | 4652 | 24 | AAZ17712 Adeno-associated v |
| 39 | 491.5 | 16.9 | 2271 | 21 | AAD00833 Adeno-associated v |
| 40 | 490 | 16.9 | 1662 | 20 | AAZ1586 Erythrovirus V9 DN |
| 41 | 490 | 16.9 | 2343 | 20 | AAZ1583 Erythrovirus V9 DN |
| 42 | 490 | 16.9 | 4677 | 16 | AAT49535 Human parvovirus g |
| 43 | 490 | 16.9 | 5028 | 20 | AAZ1580 Genomic DNA sequen |
| 44 | 486 | 16.7 | 1699 | 25 | ABZ59574 Human parvovirus B |
| 45 | 486 | 16.7 | 1699 | 25 | ABZ59577 Human parvovirus B |

ALIGNMENTS

RESULT 1

AAD00779

ID AAD00779 standard; DNA; 1605 BP.

XX AAD00779;

AC AAD00779;

DT 08-SEP-2000 (first entry)

DE Adeno-associated virus serotype 1 capsid protein VP3 DNA.

XX Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
KW cap protein; recombinant viral vector; gene delivery; gene therapy;
XX vaccine; transgene; VP3; ss.

OS Adeno associated virus serotype 1.

XX Key Location/Qualifiers

FH

FT CDS 1..1605
 FT /*tag= a
 FT /product= "vp3 protein"
 PN WO200028061-A2.
 PD 18-MAY-2000.
 XX 02-NOV-1999; 99WO-US25694.
 XX 05-NOV-1998; 98US-0107114.
 XX (UYPE-) UNIV PENNSYLVANIA.
 XX Wilson JM, Xiao W;
 XX WPI; 2000-376571/32.
 DR P-PSDB; AA71169.
 XX Novel adeno-associated virus serotype 1 polynucleotide useful for
 PT preparation of medicament for delivery of a transgene to a host -
 XX
 XX Claim 10; Page 96-99; 108pp; English.
 CC The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA
 CC which is characterised by two inverted terminal repeats (ITR) and open
 CC reading frames for rep and capsid (cap) proteins. The rep reading frame
 CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
 CC reading frame encodes three structural proteins, VP1, VP2 and VP3.
 CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap
 CC coding regions, are useful in production of recombinant viral vectors
 CC for gene delivery. These vectors can be used as gene therapy
 CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
 CC not induce the formation of neutralising antibodies specific to any
 CC serotype of AAV hence is useful for transforming host cells, and in the
 CC preparation of a medicament for the delivery of transgene to a host.
 CC The present sequence is an AAV-1 DNA encoding a cap protein vp3 which is
 CC useful in the production of recombinant viral vector for gene delivery.
 XX
 SQ Sequence 1605 BP; 431 A; 466 C; 361 G; 347 T; 0 other;

Alignment Scores:
 Pred. No.: 1,07e-268 Length: 1605
 Score: 2906.00 Matches: 534
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

US-09-807-802A-17 (1-534) x AAD000779 (1-1605)

Db 301 AACAAATGGGGATTCCGCGCCCAAGAGACTCAACTTCAAACCTCTTCAAACATCAAGTCAAG 360
 QY 121 GluValThrThrAsnAspGlyValThrThrIleAlaAsnLeuThrSerThrValGln 140
 Db 361 GAGGTCAAGCAATGATGGCGTCACCAACCATCTAATAACCTTACCAGCAGCGTTCAA 420
 QY 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
 Db 421 GTCTTCTCGGACTCGGAGTACCACTTCGCTACGTCCTCGGCTCTGGCAGCACCAGGCTGC 480
 QY 161 LeuProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
 Db 481 CTCCCTCGTTCCCGGGGACGTGTTATGATTCGCAATACGGCTACTGACGCTCAAC 540
 QY 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
 Db 541 AATGGCAGCCAAAGCGGTGGGACGTTTCATCTTTTACTGCGCTGGGAATATTTCCCTTCTCAG 600
 QY 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
 Db 601 ATGCTGAGACGGGCACACACTTTTACCTTCAGCTACACCTTTGAGGAAGTGCCTTTCCAC 660
 QY 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
 Db 661 AGCAGCTACGCGCACAGCAGAGCGCTGACCGGCTGATGAATCTCTCTCATCGACCAATAC 720
 QY 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
 Db 721 CTGTATTACCTGGAACAGAACTCAAATCAGTCCGGAAGTGCCTCAACAAACAGAGACTTGTG 780
 QY 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLysLeuProGlyPro 280
 Db 781 TTTAGCCGTGGTCTCCAGCTGGCATGCTGTTCAGCCCAAAACCTGCTACCTGGACCC 840
 QY 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
 Db 841 TGTATCGCAGCAGCGCGTTTCTAAAAACAAAAACAGACAAACAAACAGCAATTTTACC 900
 QY 301 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr 320
 Db 901 TGGACTGGTGTCTTCAAAATATTAACCTCANTGGCGGTGAATCCATCATCATCACTGGCACT 960
 QY 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
 Db 961 GCTATGGCTCACACAAAGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1020
 QY 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
 Db 1021 TTTGGAAAAGAGAGAGCGCGGAGCTTCAAAACACTGCAATTTGGCAATGTCTATGATTACAGAC 1080
 QY 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
 Db 1081 GAAGAGGAATTAAGACCACTAACCTGTGGCACCAGAAAGATTTGGGACCGCTGGCAGTTC 1140
 QY 381 AsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
 Db 1141 AATTTCCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1200
 QY 401 ProGlyMetValTrpGlnAspArgValTyrLeuGlnGlyProIleTrpAlaValIle 420
 Db 1201 CTGGCATGTGTGGCAAGATAGAGACGTGTACCTGCGAGGGTCCCAATTTGGGCGCAAAAT 1260
 QY 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
 Db 1261 CTTCAACAGATGGACACTTTTCAACCGTCTCTCTTATGGCGGCTTTGGACTCAAGAAC 1320
 QY 441 ProProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460
 Db 1321 CGGCTCTCTCAGATCTCTCAATAAAACAGCGCTGTCTCTGCGAATCTCTCGCGCGGAGTTT 1380
 QY 461 SerAlaThrLysPheAsnSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
 Db 1381 TCAGCTACAAAGTTTGTCTTCAATTCATCACTCAACCAATCTCCACAGGACAGTGTAGTGTGGAA 1440

QY 481 IleGluTrpGluLeuGlnLysGluAenSerLysArgTrpAenProGluValGlnTrpThr 500
 DB 1441 ATTGAATGGAGCTGCAGAAAGAAACAGCAGCGCTGGAATCCCGAAGTGCAGTACACA 1500
 QY 501 SerAenTyrAlaLysSerAlaAenValAapPheThrValAapAenAenGlyLeuTyrThr 520
 DB 1501 TCCAAATTATGCAAAATCTGCCAACGTTGATTTTACTGTGGACAAACAAATGGACTTTATACT 1560
 QY 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
 DB 1561 GAGCCTCGCCCATTTGGCACCCTTACCTTACCCCTGCCCTG 1602

RESULT 2

AAD00778

ID AAD00778 standard; DNA; 1800 BP.

AC AAD00778;

DB 08-SEP-2000 (first entry)

DE Adeno-associated virus serotype 1 capsid protein VP2 DNA.

KW Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
 KW cap protein; recombinant viral vector; gene delivery; gene therapy;
 KW vaccine; transgene; VP2; ss.

OS Adeno associated virus serotype 1.

FH Key Location/Qualifiers

FT CDS 1..1800

FT FT /*tag= a

FT FT /product= "VP2 protein"

FT FT /partial

XX WO200028061-A2.

XX 18-MAY-2000.

XX 02-NOV-1999; 99WO-US25694.

XX 05-NOV-1998; 98US-0107114.

XX (UYPE-) UNIV PENNSYLVANIA.

XX Wilson JM, Xiao W;

XX WPI: 2000-376571/32.

XX P-PSDB; AAY71168.

XX Novel adeno-associated virus serotype 1 polynucleotide useful for
 XX preparation of medicament for delivery of a transgene to a host -
 XX Claim 10; Page 90-93; 108pp; English.

XX The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA
 XX which is characterised by two inverted terminal repeats (ITR) and open
 XX reading frames for rep and capsid (cap) proteins. The rep reading frame
 XX encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
 XX reading frame encodes three structural proteins, VP1, VP2 and VP3.
 XX The AAV-1 sequence or its fragments particularly ITRs, rep and cap
 XX coding regions, are useful in production of recombinant viral vectors
 XX for gene delivery. These vectors can be used as gene therapy
 XX vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
 XX not induce the formation of neutralising antibodies specific to any
 XX serotype of AAV hence is useful for transforming host cells, and in the
 XX preparation of a medicament for the delivery of transgene to a host.
 XX The present sequence is an AAV-1 DNA encoding a cap protein VP2 which is
 XX useful in the production of recombinant viral vector for gene delivery.

SQ Sequence 1800 BP; 483 A; 532 C; 409 G; 376 T; 0 other;

Alignment Scores:

| | | | |
|------------------------|-----------|---------------|------|
| Pred. No.: | 1.27e-268 | Length: | 1800 |
| Score: | 2906.00 | Matches: | 534 |
| Percent Similarity: | 100.00% | Conservative: | 0 |
| Best Local Similarity: | 100.00% | Mismatches: | 0 |
| Query Match: | 100.00% | Indels: | 0 |
| DB: | 21 | Gaps: | 0 |

US-09-807-802A-17 (1-534) x AAD00778 (1-1800)

| | | | |
|----|------|--|------|
| QY | 1 | MetAlaSerGlyGlyGlyAlaProMetAlaAapAenAenGluGlyAlaAapGlyValGly | 20 |
| DB | 196 | ATGGCTTCAGGCGGTGGCGCACCAATGGCAGACAAATACGAGAGGCGCGAGCGAGTGGGT | 255 |
| QY | 21 | AsnAlaSerGlyAenTrpHisCysAapSerThrTrpLeuGlyAapArgValIleThrThr | 40 |
| DB | 256 | AATGCCTCAGGAATTTGGCATTTGCATTTCCATATGGCTGGCGCAGACAGTTCATCACCAC | 315 |
| QY | 41 | SerThrArgThrTrpAlaLeuProThrTyrAsnAenHisLeuTyrLysGlnIleSerSer | 60 |
| DB | 316 | AGCACCCGCACTTGGGCTTGGCCACCTACAAATTAACCACTTCAAGCAAAATCTCCAGT | 375 |
| QY | 61 | AlaSerThrGlyAlaSerAenAapAenHisTyrPheGlyTyrSerThrProTrpGlyTyr | 80 |
| DB | 376 | GCTTCAACGGGGCGCAGCAGCAGCACTACTTTCGGCTACAGACACCCCTGGGGGTAT | 435 |
| QY | 81 | PheAapPheAenArgPheHisCysHisPheSerProArgAapTrpGlnArgLeuIleAen | 100 |
| DB | 436 | TTTGATTTCAACAGATTTCCACTTGCCTTTTCCACCGTGCCTGCAGCGACTCATCAAC | 495 |
| QY | 101 | AsnAenTrpGlyPheArgProLysArgLeuAenPheLysLeuPheAenIleGlnValLys | 120 |
| DB | 496 | AACAAATGGGGATTTCCGGCCCCAGAGACTCAACTTCAAACTTCAACATCAAGTCAAG | 555 |
| QY | 121 | GluValThrThrAenAapGlyValThrThrIleAlaAenAenLeuThrSerThrValGln | 140 |
| DB | 556 | GAGGTCAAGCAATGATGGCTCACCAACCATCGCTAATACTTACCAGCAGCGTTCAA | 615 |
| QY | 141 | ValPheSerAapSerGlyTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys | 160 |
| DB | 616 | GTCTTCTCGGACTCGGAGTACAGCTTCCGTACGTCTCTCGGCTCTGCGCAGCGGCTGC | 675 |
| QY | 161 | LeuProProPheProAlaAapValPheMetIleProGlnTyrGlyTyrLeuThrLeuAen | 180 |
| DB | 676 | CTCCCTCCGTTCCCGGGGAGCGTGTTCATGATTCGCAATACGGCTACTGACGCTCAAC | 735 |
| QY | 181 | AsnGlySerGlnAlaValArgSerSerPheTyrCysLeuGluTyrPheProSerGln | 200 |
| DB | 736 | AATGGCAGCCGAGCGGTGGGAGCTTCATCTCTTTTACTGCTTGGAAATATTTCCCTTCTCAG | 795 |
| QY | 201 | MetLeuArgThrGlyAenAapPheThrPheSerTyrThrPheGluGluValProPheHis | 220 |
| DB | 796 | ATGCTGAGACGGGCAACAACCTTTTACCTTCAGCTACACCTTTTGAGGAAGTGCCTTTCCAC | 855 |
| QY | 221 | SerSerTyrAlaHisSerGlnSerLeuAapArgLeuMetAenProLeuIleAapGlnTyr | 240 |
| DB | 856 | AGCAGCTACGCGCACAGCCAGAGCTCGGACCGGCTGATGAATCTCTCATCGACCAATAC | 915 |
| QY | 241 | LeuTyrTyrLeuAenAapThrGlnAenGlnSerGlySerAlaGlnAenLysAapLeuLeu | 260 |
| DB | 916 | CTGTATTACCTGAACAGACTCAAAATCAGTCCGGAAGTGCCTCAAAACAAGGACTTGTG | 975 |
| QY | 261 | PheSerArgGlySerProAlaGlyMetSerValGlnProLysAenTrpLeuProGlyPro | 280 |
| DB | 976 | TTTACCGCTGGGTCTCCAGCTGGCATGTCTCTTTCAGCCCAAAAATCTGCTACCTGGACCC | 1035 |
| QY | 281 | CysTyrArgGlnGlnArgValSerLysThrLysThrAapAenAenAenAenAenPheThr | 300 |
| DB | 1036 | TGTTATCGCAGCAGCGGTTTCTTAAACAAAAACAGACAAACAAACAGCAATTTTACC | 1095 |
| QY | 301 | TrpThrGlyAlaSerLysTyrAsnLeuAenGlyArgGluSerIleIleAenProGlyThr | 320 |
| DB | 1096 | TGGACTGGTGTCTTCAAAATATAACCTCAATGGGCGTGAATCCATCATCAACCTTGGCACT | 1155 |

```
QY 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetile 340
DB 1156 GCTATGGGCTCACAAAGACGACGAAGAGTTCCTTCCCATGAGCGGTGTCATGATT 1215
QY 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetileThrAsp 360
DB 1216 TTTGGAAAGAGAGCGCGAGCTTCAAACTGCTATTGGCAATGTCTATGATTACAGAC 1275
QY 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
DB 1276 GAAGAGGAAATTAAGGCCACTAACCTGTGGCCACCGAAGATTTGGGACGTGGCAGTC 1335
QY 381 AsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyValAlaLeu 400
DB 1336 AATTTCCAGAGCAGCAGCAGCAGCCTCGCAGCGAGATGTGCATGCTATGGGACATTA 1395
QY 401 ProGlyMetValTrpGlnAspArgAspValTyLeuGlnGlyProIleTrpAlaLysile 420
DB 1396 CTGGCATGCTGGCAAGATAGAGAGCTGTACTGCAGGGTCCCATTTGGGCCCAAAATT 1455
QY 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
DB 1456 CCTCACACAGATGGACACTTTTCCCGCTCTCTTATGGGCGGCTTTGGACTCAAGAAC 1515
QY 441 ProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460
DB 1516 CCGCTTCTCAGATCTCTCATCAAAACACGCTGTCTTCCGAGATCTTCCGCGGAGTTT 1575
QY 461 SerAlaThrLysPheAlaSerPheIleThrGlnTySerThrGlyGlnValSerValGlu 480
DB 1576 TCAGCTACAAAGTTTGTCTTCATTCACCCCAATCTCCACAGGACAAAGTGTGGAA 1635
QY 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyThr 500
DB 1636 ATTGAATGGGAGCTGCAGAAAGAAACACGAGCGCTGGAATCCCGAAGTGCAGTACACA 1695
QY 501 SerAsnTyAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyThr 520
DB 1696 TCCAAATATGCAAAATCTGCCAAGTGTGATTTACTGTGGACAAACATGACTTTATACT 1755
QY 521 GluProArgProIleGlyThrArgTyLeuThrArgProLeu 534
DB 1756 GAGCCTCGCCCATTTGGCACCCTTACCTTACCCGTCCTCCCTG 1797
RESULT 3
ID AAD00777 standard; DNA; 2211 BP.
AC AAD00777;
XX
XX
DT 08-SEP-2000 (first entry)
DE
DE Adeno-associated virus serotype 1 capsid protein VP1 DNA.
XX
KW Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
KW cap protein; recombinant viral vector; gene delivery; gene therapy;
KW vaccine; transgene; VP1; ss.
XX
OS Adeno associated virus serotype 1.
XX
XX Location/Qualifiers
FH 1..2211
FT CDS
FT /tag= a
FT /product= "vp1 protein"
XX
XX WO200028061-A2.
XX
XX 18-MAY-2000.
XX
XX 02-NOV-1999; 99WO-US25694.
XX
XX 05-NOV-1998; 98US-0107114.
XX
XX
```

```
PA (UYPE-) UNIV PENNSYLVANIA.
XX
XX Wilson JM, Xiao W;
DR WPI; 2000-376571/32.
XX P-PSDB; MAY71167.
PT Novel adeno-associated virus serotype 1 polynucleotide useful for
XX preparation of medicament for delivery of a transgene to a host -
PS Claim 10; Page 83-87; 108pp; English.
XX
CC The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA
CC which is characterised by two inverted terminal repeats (ITR) and open
CC reading frames for rep and capsid (cap) proteins. The rep reading frame
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.
CC The AAV-1 sequence or its fragments particularly ITR6, rep and cap
CC coding regions, are useful in production of recombinant viral vectors
CC for gene delivery. These vectors can be used as gene therapy
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
CC not induce the formation of neutralising antibodies specific to any
CC serotype of AAV hence is useful for transforming host cells, and in the
CC preparation of a medicament for the delivery of transgene to a host.
CC The present sequence is an AAV-1 DNA encoding a cap protein VP1 which is
CC useful in the production of recombinant viral vector for gene delivery.
XX
SQ Sequence 2211 BP; 577 A; 652 C; 538 G; 444 T; 0 other;
```

Alignment Scores:

```
Pred. No.: 1.7e-268 Length: 2211
Score: 2906.00 Matches: 534
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0
```

US-09-807-802A-17 (1-534) x AAD00777 (1-2211)

```
QY 1 MetAlaSerGlyGlyValaProMetAlaAspAsnAsnGluValaAspGlyValGly 20
DB 607 ATGGCTTCAGGCGGTGGCGCCACCAATGCGACACATAACGAAAGGCCGCCGAGTGGGT 666
QY 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 40
DB 667 AATGCTCAGGAAATTTGGCATTTCCATGCTGCTGGCGCAGAGTCATCACCACC 726
QY 41 SerThrArgThrTrpAlaLeuProThrTyTrpAsnAsnHisLeuTyTyLysGlnIleSerSer 60
DB 727 AGCACCCGACCTGGGCGCTTGGCCACCTACAAATACCAACCTCTTACAAAGCAATCTCCAGT 786
QY 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyTrpPheGlyTyTrpSerThrProTrpGlyTy 80
DB 787 GCTTCAACGGGGGCCGAGCAACCAACCACTACTTTCGGCTACGACACCCCTTGGGGGTAT 846
QY 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
DB 847 TTTGATTTCAACAGATTCCACTGCCACTTTTCCACCACTGACTGGCAGGACTCATCAAC 906
QY 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 120
DB 907 AACAAATTTGGGATTCGGGCCCAAGAGACTCAACTTCAAACTCTTCAACATCCAGTCAAG 966
QY 121 GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 140
DB 967 GAGGTCAACAGCAAGATGATGGCGTCAACCACTCGTATACCTTACCTACGACGCTTCAA 1026
QY 141 ValPheSerAspSerGlyTyTrpGlnLeuProTyValLeuGlySerAlaHisGlnGlyCys 160
DB 1027 GTCTTCTCGACTCGGAGTACAGCTTCGCTACGCTCTCGGCTCTGCGCAGCAGGCTGC 1086
QY 161 LeuProProPheProAlaAspValPheMetIleProGlnTyGlyTyLeuThrLeuAsn 180
```

```
Db 1087 CTCCTCCGCTCCCGCGGACGCTGTTCATGATTCGCAATACGGCTACCTGACGCTCAAC 1146
Qy 181 AsnGlySerGlnAlaValGlyArgSerPheTyrCysLeuGluTyrPheProSerGln 200
Db 1147 AATGGCAGCCAGCGCTGGAGCTTCATCCCTTTACTGCTCCGGAATATTCCTCTTCAG 1206
Qy 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
Db 1207 ATGCTGAGACGGGCAACAACCTTTACCTTCAGCTACACCTTTGAGGAAGTGCTTTCCAC 1266
Qy 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
Db 1267 AGCAGCTACGGCAGCAGCAGCGCTGGACCGGCTGATCAATCTCTCATCGACCAATAC 1326
Qy 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
Db 1327 CTGTATTACCTGAACAGAACTCAAAATCAGTCCGGAAGTGCCCAAAACAGAGACTTGCTG 1386
Qy 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
Db 1387 TTATGCCGTGGGTCTCCAGCTGGCATGCTGTTCAGCCCAAAACTGCTACCTGGACCC 1446
Qy 281 CysTyrArgGlnClnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
Db 1447 TGTTATCGCGCAGCAGCGGCTTTCTAAACAAACAAACAGACAAACAGCAATTTTACC 1506
Qy 301 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr 320
Db 1507 TGGACTGGTGCTTCAAAATATTAACCTCAATGGCGCTGAATCATCATCATCAACCTGGCACT 1566
Qy 321 AlaMetAlaSerHisGlyAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
Db 1567 GCTATGGCTCTCACAAAGACGACGAAGCAAGTGTCTTCCCATGAGCGGTGTCATGATT 1626
Qy 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
Db 1627 TTGGAAAGAGAGCGCGGAGCTTCAAACTGCTATGGCAATGTCATGATTTACAGAC 1686
Qy 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
Db 1687 GAAGAGGAATTAAGCCACTAACTCTGTGGCCACCGAAAGATTGGGACCGTGGCAGTC 1746
Qy 381 AsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
Db 1747 AATTTCCAGCAGCAGCAGCAGACCCCTGCGACCGGAGATGTGCATGCTATGGGAGCATTA 1806
Qy 401 ProGlyMetValTrpGlnAspArgPheValTyrLeuGlnGlyProIleTrpAlaValIle 420
Db 1807 CCTGGCATGGTGGCAAGATAGACGTGTACCTGACGGGTCCCATTTGGGCCCAAAATT 1866
Qy 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
Db 1867 CCTCACAGATGGACACTTTCACCGCTCTCTCTTATGGCGGCTTTGGACTCAAGAAC 1926
Qy 441 ProProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460
Db 1927 CGCGCTCTCTCAGATCTCATCAAAACACGCTGTTCCTGCGAATCTCTCGCGGAGTTT 1986
Qy 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
Db 1987 TCAGCTACAAAGTTTCTTCATTCATCCCAATCTCCAGCACTCCACAGGACAGTGAAGTGGAA 2046
Qy 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrThr 500
Db 2047 ATTGAATGGAGCTGCAGAAAGAAACAGACAGCGCTGGAATCCCGAAGTCAGTACACA 2106
Qy 501 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 520
Db 2107 TCCAAATTATGCAAAATCTGCCAACGTTGATTTTACTGTGGACAAACAATGGACTTTATACT 2166
Qy 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
Db 2167 GAGCCTCGCCCCATTGGCAGCCGCTTACCTTACCCGTCCTCCCTG 2208
```

RESULT 4

```
AAD00772
ID AAD00772 standard; DNA; 4718 BP.
XX
AC AAD00772;
XX
DT 08-SEP-2000 (first entry)
XX
DE Adeno-associated virus serotype 1 DNA.
XX
KW Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
KW cap protein; recombinant viral vector; gene delivery; gene therapy;
KW vaccine; transgene; ss.
XX
OS Adeno associated virus serotype 1.
XX
FH Key Location/Qualifiers
FT repeat_unit 1..143
FT /tag= a
FT /label= 5' ITR
FT /note= "Inverted terminal repeat which is capable
FT of forming T-shaped hairpin structure"
FT protein_bind 89..110
FT /tag= b
FT /bound_moiety= "Rep protein"
FT misc_feature 124..125
FT /tag= c
FT /note= "Terminal resolute site (TRS)"
FT misc_binding 219..226
FT /tag= d
FT /bound_moiety= "USF"
FT /note= "E box"
FT promoter 236..299
FT /tag= e
FT /label= P5_promoter
FT protein_bind 237..245
FT /tag= f
FT /bound_moiety= "YY1 factor"
FT TATA_signal 270..275
FT /tag= g
FT /label= P5_TATA-Box
FT misc_feature 299..306
FT /tag= h
FT /note= "YY1/p5 RNA"
FT CDS 335..2206
FT /tag= i
FT /product= "Rep 78"
FT /function= "regulates replication and integration
FT of AAV DNA into host cell's chromosome"
FT CDS 335..2272
FT /tag= j
FT /product= "Rep 68"
FT /function= "regulates replication and integration
FT of AAV DNA into host cell's chromosome"
FT /note= "The coding region is interrupted by intron"
FT intron 1924..2220
FT /tag= k
FT /note= "This region interrupts the coding sequence
FT of Rep 68 and Rep 40"
FT TATA_signal 857..862
FT /tag= l
FT /label= P19_TATA_Box
FT misc_feature 882..883
FT /tag= m
FT /note= "P19 RNA"
FT CDS 1007..2206
FT /tag= n
FT /product= "Rep 52"
FT /function= "regulates replication and integration
FT of AAV DNA into host cell's chromosome"
FT CDS 1007..2272
FT /tag= o
```

FT /product= "Rep 40"
 FT /function= "regulates replication and integration
 FT of AAV DNA into host cell's chromosome"
 FT /note= "The coding region is interrupted by intron"
 FT 1836..1841
 FT /*tag= p
 FT TATA_signal
 FT /label= P40_TATA-BOX
 FT 1875..1876
 FT misc_feature
 FT /*tag= q
 FT /note= "P40 RNA"
 FT 2223..4433
 FT CDS
 FT /*tag= x
 FT /product= "VP1 protein"
 FT /note= "Capsid protein"
 FT 2634..4433
 FT CDS
 FT /*tag= s
 FT /product= "VP2 protein"
 FT /note= "Capsid protein"
 FT /partial
 FT 2829..4433
 FT CDS
 FT /*tag= t
 FT /product= "VP3 protein"
 FT /note= "Capsid protein"
 FT 4447..4452
 FT polyA_signal
 FT /*tag= u
 FT 4576..4718
 FT repeat_unit
 FT /*tag= v
 FT /label= 3'_ITR
 FT /note= "Inverted terminal repeat which is capable
 FT of forming T-shaped hairpin structure"
 FT
 XX WO200028061-A2.

XX
 PD 18-MAY-2000.

XX
 PF 02-NOV-1999; 99WO-US25694.

XX
 PR 05-NOV-1998; 98US-0107114.

XX
 PA (UYPB-) UNIV PENNSYLVANIA.

XX
 PI Wilson JM, Xiao W;

XX
 DR WPI: 2000-376571/32.
 DR P-PSDB; AAY71161, AAY71164, AAY71165, AAY71166, AAY71167, AAY71168,
 DR AAY71169.

XX
 PT Novel adeno-associated virus serotype 1 polynucleotide useful for
 PT preparation of medicament for delivery of a transgene to a host -

XX
 PS Claim 1; Fig 1; 109pp; English.

XX
 CC The present sequence is an adeno-associated virus serotype 1 (AAV-1)
 CC DNA characterised by two inverted terminal repeats (ITR) and open
 CC reading frames for rep and capsid (cap) proteins. The rep reading frame
 CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
 CC reading frame encodes three structural proteins, VP1, VP2 and VP3.
 CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap
 CC coding regions, are useful in production of recombinant viral vectors
 CC for gene delivery. These vectors can be used as gene therapy
 CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
 CC not induce the formation of neutralising antibodies specific to any
 CC serotype of AAV hence is useful for transforming host cells, and in the
 CC preparation of a medicament for the delivery of transgene to a host.

XX
 SQ Sequence 4718 BP; 1121 A; 1393 C; 1273 G; 931 T; 0 other;

Alignment Scores:

| | | | |
|------------------------|-----------|---------------|------|
| Pred. No.: | 5,03e-268 | Length: | 4718 |
| Score: | 2906.00 | Matches: | 534 |
| Percent Similarity: | 100.00% | Conservative: | 0 |
| Best Local Similarity: | 100.00% | Mismatches: | 0 |
| Query Match: | 100.00% | Indels: | 0 |

| | | | |
|---|------|---|------|
| DB: | 21 | Gaps: | 0 |
| US-09-807-802A-17 (1-534) x AAD00772 (1-4718) | | | |
| QY | 1 | MetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGly | 20 |
| DB | 2829 | ATGGCTTCAGGCGGTGGCGCACCAATACGACACAATACGAAGCGCGCAGCGAGTGGGT | 2888 |
| QY | 21 | AsnAlaSerGlyAenTtpHieCyAspSerThrTrpLeuGlyAspAigValIleThr | 40 |
| DB | 2889 | AATGCCTCAGGAAATTTGGCATTTCCATATGGCTGGCGGACAGAGTATATCACCACC | 2948 |
| QY | 41 | SerThrArgThrTtpAlaLeuProThrTyrAsnAsnHieLeuTyrLysGlnIleSerSer | 60 |
| DB | 2949 | AGCACCGGCACCTGGGCTTCCACCTATCAATAACACCTCTTCAACAGCAAAATCTCCAGT | 3008 |
| QY | 61 | AlaSerThrGlyAlaSerAsnAspAsnHieTyrPheGlyTyrSerThrProTtpGlyTyr | 80 |
| DB | 3009 | GCTTCAACGGGGGCGCAGCAACACCTACTTCGGCTACAGCACCCCTGGGGGTAT | 3068 |
| QY | 81 | PheAspPheAsnArgPheHieCyAspHiePheSerProArgAspTtpGlnArgLeuIleAsn | 100 |
| DB | 3069 | TTTGATTTCAACAGATTCCTGCTGCTTTTCCACACGTGACTGGCAGCGACTCATCAAC | 3128 |
| QY | 101 | AsnAsnTtpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys | 120 |
| DB | 3129 | AACAATTCGGGATTCGGGCGCCCAAGAGACTCAACTTCAAACTCTTCAACATCCACGCTCAAG | 3188 |
| QY | 121 | GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln | 140 |
| DB | 3189 | GAGGTCAACGCAATGATGGCGTCAACACCTCGCTAATACCTTACCAGCACGGTTCAA | 3248 |
| QY | 141 | ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHieGlnGlyCys | 160 |
| DB | 3249 | GTCTTCTCGACTCGAGTACAGCTTCGCTACGCTCTCGGCTCTGGCAGCACGGGCTGC | 3308 |
| QY | 161 | LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn | 180 |
| DB | 3309 | CTCCCTCCGTTCCGGGCGGACGTGTTTCATGATTCGGCAATACGGCTACTGACGCTCAAC | 3368 |
| QY | 181 | AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln | 200 |
| DB | 3369 | AATGGCAGCCAAAGCGTGGGAGCTTCATCTCTTACTGCTCGGAATATTTCTCTCTCAG | 3428 |
| QY | 201 | MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValIleProPheHie | 220 |
| DB | 3429 | ATGCTGAGACGGGCAACAACCTTACCTTCAGCTACACTTTGAGGAGAGTGGCTTCCAC | 3488 |
| QY | 221 | SerSerTyrAlaHieSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr | 240 |
| DB | 3489 | AGCAGCTACGGCACAGCCAGAGCTGGACCGGCTGATGAATCTCTCATCGACCAATAC | 3548 |
| QY | 241 | LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysLeuLeu | 260 |
| DB | 3549 | CTGTATTACCTGAACAGAACTCAAAATCAGTCCGGAAGTCCCAAAACAGAGACTTGCTG | 3608 |
| QY | 261 | PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro | 280 |
| DB | 3609 | TTTACCGTGGGTCTCCAGCTGGCATGCTGTTCAGCCCAAAAACCTGGCTACTGGACCC | 3668 |
| QY | 281 | CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr | 300 |
| DB | 3669 | TGTTATCGGCAGCAGCGGCTTTCTAAAAAACAACAGACACACACACACAAATTTTACC | 3728 |
| QY | 301 | TtpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr | 320 |
| DB | 3729 | TGGACTGGTGTCTCAAAATATTAACCTCAATGGGCGTGAATCTCATCATCAACCTCGCACT | 3788 |
| QY | 321 | AlaMetAlaSerHieLysAspAspGluAspLysPhePheProMetSerGlyValMetIle | 340 |
| DB | 3789 | GCTATGGCTTCACACAAAGACGACGAGACAAGTCTTTTCCCATGAGCGGTGTCATGATT | 3848 |
| QY | 341 | PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAsnValMetIleThrAsp | 360 |

```

Db 3849 TTTGGAAGAGAGCGCGGAGCTTCAAACTGCTGATGGCAATGTGATGATACAGAC 3908
Qy 361 GluGluGluIleLeuAlaThrAsnProValAlaThrGluAtpGheGlyThrValAlaVal 380
Db 3909 GAAAGAGAAATTAAGCCACTAACCTGTGGCCACCGAAGATTTGGACCGTGGCAGTC 3969
Qy 381 AsnGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
Db 3969 AATTTCCAGAGCAGCAGCACACACCTGCGACCGGAGATGTGATGCTATGGGACATTA 4028
Qy 401 ProGlyMetValTrpGlnAspArgValTrpLeuGlnGlyProIleTrpAlaLysIle 420
Db 4029 CCGTGCATGTGTGGCAAGATAGAGACGTGTACTTCCGAGGGTCCCAATTTGGGCCAAAT 4088
Qy 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
Db 4089 CTTACACAGATGGACATTTTCCCGCTCTCTTTATGGCGCGCTTTGGACTCAAGAAC 4148
Qy 441 ProProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460
Db 4149 CCGCTCTCTCAGATCTCATCAAAACAGCGCTGTCTCTGCGAATCTCCGCGGAGTTT 4208
Qy 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
Db 4209 TCAGCTACAAAGTTTGTCTTCAATTCATCACCAATATCTCCACAGGACAAAGTGAGTGGAA 4268
Qy 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrThr 500
Db 4269 ATTGAATGGAGCTGCAGAAAGAAACAGACAGCGCTGGAAATCCCGAAGTGCAGTACAA 4328
Qy 501 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 520
Db 4329 TCCAAATATGCAAAATCTGCCAACGTTGATTTTACTGTGGACAAACAATGGACTTTTACT 4388
Qy 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
Db 4389 GAGCCTCGCCCATTTGGCACCGGTTACCTTACCCCGTCCCGCTG 4430

RESULT 5
AAF23749
ID AAF23749 standard; DNA; 4683 BP.
XX
AC AAF23749;
XX
DT 28-MAR-2001 (first entry)
XX
DE AAV6 DNA sequence.
XX
KW AAV6; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
KW athrosclerosis; sickle cell anaemia; thalassaemia;
KW blood clotting disorder; diabetes; ss.
XX
OS Adeno associated virus.
XX
FN US6156303-A.
XX
PD 05-DEC-2000.
XX
PF 11-JUN-1997; 97US-0873168.
XX
PR 11-JUN-1997; 97US-0873168.
XX
PA (UNIW ) UNIV WASHINGTON.
XX
PI Russell DW, Rutledge EA;
XX
DR WPI; 2001-060164/07.
XX
Adeno-associated virus serotype 6 and viral vector derived from it for
PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
PT syndrome, sickle cell anemia, thalassaemia and diabetes
XX

```

```

PS Claim 1; Fig 1; 50pp; English.
XX
CC The present invention relates to adeno-associated virus serotypes. The
CC present sequence is the DNA sequence of one such serotype (AAV6). AAV6
CC can be used to construct AAV viral vectors for use in gene therapy for a
CC range of disorders: cystic fibrosis, cancer, AIDS, athrosclerosis,
CC sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.
CC The AAV viral vectors have increased transduction efficiency of a
CC particular host cell as the AAV virion containing the AAV vector genome
CC can be modified to express a capsid protein of an AAV serotype that
CC transduces the selected host cell.
XX
SQ Sequence 4683 BP; 1114 A; 1363 C; 1277 G; 929 T; 0 other;

Alignment Scores:
Pred. No.: 6.41e-266 Length: 4683
Score: 2884.00 Matches: 529
Percent Similarity: 99.63% Conservative: 3
Best Local Similarity: 99.06% Mismatches: 2
Query Match: 99.24% Indels: 0
DB: 22 Gaps: 0

US-09-807-802A-17 (1-534) x AAF23749 (1-4683)
Qy 1 MetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGly 20
Db 2814 ATGGCTTCAGCGGTGGCGCACCAATGCGACACAATTAACGAAGCGCGCAGCGAGTGGGT 2873
Qy 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 40
Db 2874 AATGCCCTCAGGAATTTGGCATTTCCATGTGGCTGGCGGACAGAGTTCATCCACC 2933
Qy 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSer 60
Db 2934 AGCACCCGAAATCGGCCCTTCCACCTATTAACCAACCACTTCAACGAAATCTCCAGT 2993
Qy 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
Db 2994 GCTTCAACGGGGGCGCAGCAGACACCACTACTTTCGGCTACAGCACCCCTCGGGGGTAT 3053
Qy 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
Db 3054 TTTGATTTCAACAGATTCCACTGCGCATTTCTCACACGTGACTGGCAGCGACTCATCAAC 3113
Qy 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 120
Db 3114 AACATTTGGGATTCGGGCCCAAGAGACTCAACTCAAGCTCTTCAACATCCAAGTCAAG 3173
Qy 121 GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 140
Db 3174 GAGGTCAAGCAATGATGGCGTCACGACCATCGCTAATAACCTTACCAGCAGCGTTCAA 3233
Qy 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
Db 3234 GTCTTCTCGGACTCGGAGTACAGTTGCGGTACGTCTTCGGCTCTGGCCACAGGGGCTGC 3293
Qy 161 LeuProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
Db 3294 CTCCCTCCGTTCCCGGGGACGTGTTCATGATTCCGCGAGTACGGCTACTTAACGCTCAAC 3353
Qy 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
Db 3354 AATGGCAGCCAGGCGAGTGGGACGGTCTATCTCTTTTACTGCTCCCTGGGAATATTTCCCATCGGAG 3413
Qy 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
Db 3414 ATGCTGAGAACGGGCAATTAACCTTACCTTACGCTTACACCTTCGAGGAGCTGCTCTTCCAC 3473
Qy 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
Db 3474 AGCAGCTACGCGCACAGCAGAGCGTGGACCGGCTGATGAATCTCTCTCATCGACCACTAC 3533
Qy 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260

```

Db 3534 CTGATTATCTGAACAGAACTCAGATCATCTCGGAAGTGCACAAACAGGACTTGCTG 3593

Qy 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280

Db 3594 TTTAGCCGGGGCTCTCCAGCTGGCATGTCTGTTACGCCCAAACTGGCTACCTGGACCC 3653

Qy 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsnPhenThr 300

Db 3654 TGTTACCGGCACAGCGCGTTTCTAAACAAACAAACAGACAAACAAACAGCAACTTTACC 3713

Qy 301 TrpThrGlyValAsnSerLysTyrAsnLeuAsnGlyArgGluSerIleLeuAsnProGlyThr 320

Db 3714 TGGACTGGTGCTTCAAAATATTAACCTTAATGGCGGTGATCTATATCAACCTGGCACT 3773

Qy 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340

Db 3774 GCTATGGCTTCACAAAGACGACAAAGCAAGTCTTTCCCATGAGCGGTGTCATGATT 3833

Qy 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360

Db 3834 TTTGGAAAGGAGCGCGGAGCTTCAAAACACTGCATTGGCAATGTCTATGATCACAGAC 3893

Qy 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380

Db 3894 GAAGAGGAATCAAGCCACTAACCCCGTGGCCACCGAAGATTTGGGACTGTGGCAGTC 3953

Qy 381 AsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400

Db 3954 AATCTCCAGAGCAGCAGCAGACACCCCTGGCAGCGAGATGTGATGTTATGGGAGCCTTA 4013

Qy 401 ProGlyMetValTrpGlnAspArgAspValTrpLeuGlnGlyProIleTrpAlaValIle 420

Db 4014 CTTGGAATGTGTGGCAACAGACAGACGATATACCTCGAGGGTCTTATTTGGGCCCAAAAT 4073

Qy 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440

Db 4074 CCTCACAGGATGGACACTTTCACCCGCTCTCTCTCATGGCGGCTTTGGACTTAAGCAC 4133

Qy 441 ProProGlnIleLeuLysAsnThrProValProAlaAsnProProAlaGluPhe 460

Db 4134 CGCCCTCTCAGATCTCATCAAAACACAGCCCTGTTCTCGCAATCTCTCCGCGCAGAGTTT 4193

Qy 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480

Db 4194 TCGGCTACAAAGTTTGCTTCATTCATCCAGTATTCACAGGACAGTGGCGTGGAG 4253

Qy 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrThr 500

Db 4254 ATTGAATGGAGCTGCAGAAAGAAACAGCAACAGCTGGAATCCCGAAGTGCAGTATACA 4313

Qy 501 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 520

Db 4314 TCTAACTATGCAAAATCTGCACACGTTGATTTTCACTGTGGACAAACAAATGGACTTTATCT 4373

Qy 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534

Db 4374 GAGCCTCGCCCCATTGGCACCCGCTTACCTCACCCGTCCTCTG 4415

RESULT 6

AAF23748

ID AAF23748 standard; DNA; 4722 BP.

XX

AC AAF23748;

XX

28-MAR-2001 (first entry)

XX

AAV3B DNA sequence.

XX

AAV3B; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;

XX atherosclerosis; sickle cell anemia; thalassemia;

XX blood clotting disorder; diabetes; ss.

XX

Adeno associated virus.

XX US6156303-A.

XX

XX 05-DEC-2000.

XX

XX 11-JUN-1997; 97US-0873168.

XX

XX 11-JUN-1997; 97US-0873168.

XX (UNIW) UNIV WASHINGTON.

XX Russell DW, Rutledge EA;

XX WPI; 2001-060164/07.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Alignment Scores:

Pred. No.: 5,86e-232 Length: 4722

Score: 2530.00 Matches: 460

Percent Similarity: 91.03% Conservative: 27

Best Local Similarity: 85.98% Mismatches: 46

Query Match: 87.06% Indels: 2

DB: 22 Gaps: 2

US-09-807-802A-17 (1-534) x AAF23748 (1-4722)

Qy 1 MetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGly 20

Db 2814 ATGGCTTCAGGCGGTGGCGCCACCATGGCATTTCCCAATGGCTGGGCGACAGAGTCATCACCACC 2873

Qy 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 40

Db 2874 AATTCCTCAGAAATTTGGCATTTGGCATTTCCCAATGGCTGGGCGACAGAGTCATCACCACC 2933

Qy 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSer 60

Db 2934 AGCACCAGAACCTGGGGCTCCCACTTTACAAACCAACCACTCTCTACAGCAATCTCCAGC 2993

Qy 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 80

Db 2994 CAACTCA---GGAGCTTCAACAGACACCACTACTTTGGCTACAGACCCCTTGGGGGTAT 3050

Qy 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100

Db 3051 TTTGACTTTAAACAGATTCACCTGGCCACTTCTCACCGCTGACTGGCAGCAGCTCATTAAC 3110

Qy 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 120

Db 3111 AACAACTGGGGATTCGGGCCCAAGAACTCAAGCTCTCAAGCTCTTCAACATCCAGTTAA 3170

Qy 121 GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 140

Db 3171 GAGGTCAGCAGACAGATGGCAGCAGCTATTGGCAATACCTTACCAGCAGCGTTCAA 3230

Qy 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160

Adeno-associated virus serotype 6 and viral vector derived from it for gene therapy of cystic fibrosis, cancer, acquired immunodeficiency syndrome, sickle cell anemia, thalassemia and diabetes -

Example 2; Fig 1; 50pp; English.

The present invention relates to adeno-associated virus serotypes. The present sequence is the DNA sequence of one such serotype (AAV3B). AAV3B can be used to construct AAV viral vectors for use in gene therapy for a range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis, sickle cell anaemia, thalassemia, blood clotting disorders and diabetes. The AAV viral vectors have increased transduction efficiency of a particular host cell as the AAV virion containing the AAV vector genome can be modified to express a capsid protein of an AAV serotype that transduces the selected host cell.

Sequence 4722 BP; 1204 A; 1297 C; 1243 G; 978 T; 0 other;

Db 3231 GTGTTACGGAGTATCAGCTCCGTCAGCTCGGTCGGCGACCAAGGCTGT 3290
Qy 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
Db 3291 CTCGCCCGTTTCAGCGGAGCTTTCATGGTCCCTCAGTAGTGAATCCTCACCTGAC 3350
Qy 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
Db 3351 AACGGAAGTCAAGCGGTGGAGCGCTCATCTTTTACTGCTGGAGTAGTCTCCCTTCGAG 3410
Qy 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
Db 3411 ATGCTAAGGAGCTGGAATAAATTCATATCAGTATACCTTCGAGGATGATCTTTTCAC 3470
Qy 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
Db 3471 AGCAGCTACGCTCACGCGCAGAGTTGGATCGCTTGATGAATCCTCTTATTGATCAGTAT 3530
Qy 241 LeuTyrTyrLeuAsnArgThrGln---AsnGlnSerGlySerAlaGlnAsnLysAspLeu 259
Db 3531 CTGTACTACCTGAACAGACGACGAGCAACACCTCTGGACCAACCAATCAGGCTG 3590
Qy 260 LeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTyrLeuProGly 279
Db 3591 CTTTTTAGCGGCTGGGCTCAGTCTATGCTCTTTCGACGGCCAGAAATGGCTACCTGGG 3650
Qy 280 ProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsnPhe 299
Db 3651 CCCTCTACCGGCAACAGAGACTTTCAGAGACTGCTAACGACACACACACAGTAACCTT 3710
Qy 300 ThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGly 319
Db 3711 CCTTGGACAGCGGCCAGCAATATCATCTCAATGGCCGCGACTCGTGTGAATCAGGA 3770
Qy 320 ThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMet 339
Db 3771 CCAGCTATGGCCAGTCACAGGACGATGAAGAAAATTTTCCCTATCAGCGGCAATCTA 3830
Qy 340 IlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThr 359
Db 3831 ATATTTGGCAAGAGGGACACGCGAGTAACGAGAGATTAGATTAATGATGATTACG 3890
Qy 360 AspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAla 379
Db 3891 GATGAAGAAGAGATTTCGTACCACTTCTGTGGCAACAGACAGCAGTATGGAAGTGGCA 3950
Qy 380 ValAsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAla 399
Db 3951 AATAAATTTGCAGAGCTCAAAATACAGTCCACGACTAGAACTGTCAATGATCAGGGGCC 4010
Qy 400 LeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLys 419
Db 4011 TTACTTGGCATGGTGGCAAGATCGTAGCTGTACCTTCAAGGACCTATCTGGGCAAG 4070
Qy 420 IleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLys 439
Db 4071 ATTCTCTCACCGGATGGACACTTTTCATCTCTCTCTGATGGAGGCTTTTGGACTGAAA 4130
Qy 440 AsnProProGlnIleLeuIleLysAsnThrProValProAlaAsnProAlaGlu 459
Db 4131 CATCCGCTCTCTCAATCATGATCAAAAATACTCCGGTACCGGCAATCTCTCCGAGACT 4190
Qy 460 PheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerVal 479
Db 4191 TTCAGCCCGCCCAAGTTTCTTCATTTATTCATCTAGTACTCTCAGTGGACAGTTCAGGTG 4250
Qy 480 GluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTyrAsnProGluValGlnTyr 499
Db 4251 GAAATTGAGTGGGAGCTACAGAAAGAAAACAGCAACGTTGGAATCCAGAGATTCACTGAT 4310
Qy 500 ThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyr 519

Db 4311 ACTTCCAACTACCAACAGTCTGTAATGTGGACTTTACTGTAGACACTAATGGTGTAT 4370
Qy 520 ThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
Db 4371 AGTGAACCTCGCCTATTGGAAACCGGTATCTCACAGAAACTTG 4415
RESULT 7
AAD00834
ID AAD00834 standard; DNA; 8178 BP.
XX AAD00834;
AC AAD00834;
XX 08-SEP-2000 (first entry)
DT Chimeric adeno-associated virus 2/3 capsid encoding helper plasmid.
XX Adeno-associated virus; AAV2; AAV3; chimeric; gene therapy; plasmid;
KW recombinant parvoviral vector; cellular tropism; rep gene; cap protein;
KW capsid; gene delivery; VP1; VP2; VP3; ss.
XX Chimeric - Adeno associated virus serotype 2.
OS Chimeric - Adeno associated virus serotype 3.
XX
FH Key Location/Qualifiers
CDS 251..2180
FT /*tag= a
FT /product= "Rep 68"
FT /note= "The specification states that this region
FT encodes Rep 68 protein of AAV2. However, the stop
FT position given in the specification for this CDS does
FT not correspond to stop codon"
FT 251..2114
FT /*tag= b
FT /product= "Rep 78"
FT /note= "The specification states that this region
FT encodes Rep 78 protein of AAV2. However, the stop
FT position given in the specification for this CDS does
FT not correspond to stop codon"
FT 923..2114
FT /*tag= c
FT /product= "Rep 52"
FT /note= "The specification states that this region
FT encodes Rep 52 protein of AAV2. However, the stop
FT position given in the specification for this CDS does
FT not correspond to stop codon"
FT 923..2180
FT /*tag= d
FT /product= "Rep 40"
FT /note= "The specification states that this region
FT encodes Rep 40 protein of AAV2. However, the stop
FT position given in the specification for this CDS does
FT not correspond to stop codon"
FT 2133..4342
FT /*tag= e
FT /product= "VP1 cap protein"
FT /note= "The specification states that this region
FT encodes VP1 protein of AAV2. However, the stop
FT position given in the specification for this CDS does
FT not correspond to stop codon"
FT 2544..4342
FT /*tag= f
FT /product= "VP2 cap protein"
FT /note= "The specification states that this region
FT encodes VP2 protein of AAV2. However, the positions
FT given in the specification for this CDS do not correspond
FT to start or stop codons"
FT 2739..4342
FT /*tag= g
FT /product= "VP3 cap protein"
FT /note= "The specification states that this region
FT encodes VP3 protein of AAV2 in which loops 2-4 are
FT replaced with the corresponding region from AAV3.
FT However, the stop position given in the specification
FT

FT misc_feature for this CDS does not correspond to stop codon"
 FT 3184..4092 /*tag= h
 FT /note= "AAV3 sequence"

PN WO200028004-A1.

XX 18-MAY-2000.

XX 10-NOV-1999; 99WO-US26505.

XX 10-NOV-1998; 98US-0107840.

PR 10-MAR-1999; 99US-0123651.

XX (UYNC-) UNIV NORTH CAROLINA.

XX Rabinowitz JE, Samulski RJ, Xiao W;

XX WPI; 2000-376523/32.

XX Recombinant parvoviral vectors with altered packaging, tropisms and
 PT immunogenic properties, useful in gene therapy protocols -

XX Example 27; Page 143-147; 153pp; English.

XX The patent discloses modified parvovirus vectors with advantageous
 CC antigenic properties, packaging capabilities and cellular tropisms.
 CC These vectors can be used in standard recombinant DNA protocols e.g. gene
 CC therapy for delivering nucleic acids to cells. The present sequence is
 CC a helper plasmid encoding a chimeric adeno-associated virus serotype 2/3
 CC (AAV2/3) capsid. The sequence contains the AAV2 rep coding sequences,
 CC most of the AAV2 capsid coding sequences with the exceptions that
 CC loops 2-4 from the AAV2 VP3 subunit is replaced with the corresponding
 CC region from AAV3, in a pluscript backbone. The rep
 CC coding sequence encodes four proteins responsible for replication,
 CC Rep 68, Rep 78, Rep 82 and Rep 40, and the capsid (cap) coding region
 CC encodes three structural proteins VP1, VP2 and VP3.
 CC Recombinant parvovirus comprising the chimeric capsid is useful
 CC for gene delivery.

XX Sequence 8178 BP; 2082 A; 2078 C; 2027 G; 1991 T; 0 other;

Alignment Scores:

Pred. No.: 1-82e-230 Length: 8178
 Score: 2518.00 Matches: 458
 Percent Similarity: 90.65% Conservative: 27
 Best Local Similarity: 85.61% Mismatches: 48
 Query Match: 86.65% Indels: 2
 DB: 21 Gaps: 2

US-09-807-802A-17 (1-534) x AAD00834 (1-8178)

QY 1 MetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGly 20
 Db 2739 ATGGCTACAGGAGTGGCGCACCACCAATGGCAGACAATAACGAGGCGCGCAGAGTGGGT 2798
 QY 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTripleGlyAspArgValIleThrThr 40
 Db 2799 AATTCTCCCGAATTTGGCATTGGCATTCACATGGATGGGCGACAGAGTCATCACACC 2858
 QY 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSer 60
 Db 2859 AGCACCGGAACCTGGGCGCTGCCACCTACACACACCCNCTCTACAAACAAATTTCCAGC 2918
 QY 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
 Db 2919 CAATCA---GGAGCCTCGAACGACCAATCACTACTTTGGCTACAGCACCCCTTGGGGGTAT 2975
 QY 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
 Db 2976 TTTGACTTCAACAGATTCCACTGCCACTTTTCCACCGTACGTGGCAAGACATCATCAAC 3035
 QY 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 120

Db 3036 AACAACTGGGATTCGACCCAGAGACTCACTTCAAGCTCTTTAATCATTCAGTCAA 3095
 QY 121 GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 140
 Db 3096 GAGGTCAACGAGATGACGGTACGACGAGATTGCCAATAACCTTACACGACACGGTTCCAG 3155
 QY 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
 Db 3156 GTGTTTACTGACTCGGAGTACCACTCCGTAAGTGTCTCGGTCCGGCCACCAAGGCTGT 3215
 QY 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
 Db 3216 CTCCGCGCTTCCAGCGGAGCTCTTCATGTGCTCCTCAGTATGATACCTCACCTGAAC 3275
 QY 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
 Db 3276 AACGAAATCAAGCGTGGGACGCTCATCTTTTACTGCTGGAGTACTTCCCTTCGCGAG 3335
 QY 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
 Db 3336 ATGCTAAGGACTGGAATAAATCTTCAATTCAGCTATACCTTCGAGGATGTTACCTTTTCAC 3395
 QY 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
 Db 3396 AGCAGCTACGCTCACAGCCAGAGTTTGGATCGCTTGATGAATCTCTTATTGATCAGTAT 3455
 QY 241 LeuTyrTyrLeuAsnArgThrGln---AsnGlnSerGlySerAlaGlnAsnLysAspLeu 259
 Db 3456 CTGTACTACTCTGAACAGAACCAAGAACCAACCTCTCGGAACCAACCAACCAATACACGGT 3515
 QY 260 LeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGly 279
 Db 3516 CTTTATGACAGGTGGGCTCAGTCTATGCTTTGTCAGGCGCAGAAATTTGGCTACTGGG 3575
 QY 280 ProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPhe 299
 Db 3576 CCCTGTACCAGCAACAGAGACTTTCAAAGACTGTCAAACGACTGTAAACGACCAACCAACAGTAACTTT 3635
 QY 300 ThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGly 319
 Db 3636 CTTTGGACAGCGCGCCAGCAATATATCATCTCAATGGCCGCGACTCTCGCTGATCATCAGAA 3695
 QY 320 ThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMet 339
 Db 3696 CCAGCTATGGCCAGTCCACAGGACGATGAAGAAATTTTCCCTATTCACCGCAATCTA 3755
 QY 340 IlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThr 359
 Db 3756 ATATTTGGCAAGAGGAGCAACGCGCAAGTAACGCGAGAAATTAGATAATGTATGATTACG 3815
 QY 360 AspGluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAla 379
 Db 3816 GATGAAGAGAGATTCTGACCAACCAATCTGTGGCAACAGACGATGGAATCTGGCA 3875
 QY 380 ValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValIleAlaMetGlyAla 399
 Db 3876 AATACTTCCAGAGCTCAATATACACTCCACGACTGGAACCTGTCAATCATCAGGGGCGC 3935
 QY 400 LeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLys 419
 Db 3936 TTACCTGGCATGCTGGCAAGATCGTGACGTGTACCTTCAAGGACCTATCTGGGCAAG 3995
 QY 420 IleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLys 439
 Db 3996 ATTCTCTCACCGGATGGACACTTTTCATCTCTCTGATGGGAGGCTTTGGACTGAAA 4055
 QY 440 AsnProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGlu 459
 Db 4056 CATCCGCTCTCTCAATCATGATCAAAAATACTCCGGTACCTCGGAATCTCTCGACCACC 4115
 QY 460 PheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerVal 479

Db 4116 TTCAGTGGCGCAAGTTTCTTCCTTCATCACACAGTACTCCACGGGACAGGTTCAGCGTG 4175
 Qy 480 GluileGluTTPGluLeuGlnLysGluAenSerLysArgTTPAsnProGluValGlnTyr 499
 Db 4176 GAGATCGAGTGGAGCTGCAGAGGAAACACAAACGCTGGATCCCGAATTCAGTAC 4235
 Qy 500 ThrSerAsnTyrAlaLysSerAlaAenValAspPheThrValAspAsnAsnGlyLeuTyr 519
 Db 4236 ACTTCCAACTACACAAAGTCTGTAAATCGTGGACTTACCGTGATTAATGGCTGTAT 4295
 Qy 520 ThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
 Db 4296 TCAGAGCTCGCCCATTTGGCCACAGATACCTGACTCGTAATCTG 4340

RESULT 8

AAH26327

ID AAH26327 standard; DNA; 4072 BP.

XX

AC AAH26327;

XX 02-OCT-2001 (first entry)

DT Adenovirus helper Ad cap2.

XX AAV; vector; adenovirus; helper virus; Ad cap2; loxP site;

KW gene therapy; ds.

XX Chimeric - Mastadenovirus.

OS Chimeric - Adeno associated virus.

OS Chimeric - Human cytomegalovirus.

XX WO200155361-A2.

XX 02-AUG-2001.

XX 26-JAN-2001; 2001WO-US02709.

XX 26-JAN-2000; 2000US-0178536.

XX (CHIR) CHIRON CORP.

XX Hardy SF;

XX WPI; 2001-483239/52.

XX Producing recombinant adeno-associated virus (rAAV) vector, by stably

XX infecting eukaryotic host cell with rAAV vectors, helper adenovirus or

XX herpes virus, recombinant adenovirus or herpes vectors -

XX Example 2; Page 58-59; 63pp; English.

XX The present sequence is that of adenovirus helper Ad cap2, made

XX by Cre/lox recombination and expressing adeno-associated virus

XX (AAV) VP1, 2,3 from the cytomegalovirus immediate early promoter

XX of pAdlox, and containing a loxP site. Ad cap2 was used to

XX demonstrate recombinant AAV (rAAV) production from virally

XX transduced cells. The invention provides methods and compositions

XX for producing rAAV vector particles by: (a) introducing into a host

XX cell (i) AAV packaging plasmid pFloxAAV (see AAH26326), (ii) a

XX recombinant viral vector encoding plasmid, and (iii) a plasmid

XX encoding herpesvirus, cytomegalovirus or adenoviral functions, or

XX a herpesvirus, cytomegalovirus or adenovirus itself, to produce

XX a flox AAV particles and rAAV particles; and (b) introducing into a

XX second host cell (i) the rAAV particles or (a), (ii) a vector

XX that directs expression of Cre, and (iii) a vector which directs

XX expression of herpes virus, cytomegalovirus or adenovirus helper

XX functions, such that rAAV vector particles are produced. The

XX vectors are useful for in vivo or in vitro gene therapy and also

XX for in vitro recombinant protein production.

XX Sequence 4072 BP; 1073 A; 1025 C; 992 G; 982 T; 0 other;

XX Alignment Scores:

Pred. No.: 2,13e-227 Length: 4072
 Score: 2481.50 Matches: 445
 Percent Similarity: 90.26% Conservative: 37
 Best Local Similarity: 83.33% Mismatches: 51
 Query Match: 95.39% Indels: 1
 DB: 22 Gaps: 1

US-09-807-802A-17 (1-534) x AAH26327 (1-4072)

Qy 1 MetAlaSerGlyGlyGlyAlaProMetAlaAspAenAenGluGlyAlaAspGlyValGly 20
 Db 2090 ATGGCTACAGCAGTGGCGACCAATGGCAGCAATTAACGAGGGCGCGACGGAGTGGT 2149
 Qy 21 AsnAlaSerGlyAenTTPHisCysAspSerThrTTPLeuGlyAspArgValIleThrThr 40
 Db 2150 AATTCCTCGGAAATTCGCATTGGCATTCACATGGATGGCGCAGAGTCATCACCACC 2209
 Qy 41 SerThrArgThrTTPAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSer 60
 Db 2210 AGCACCCGAACTGGGGCCCTGCCCACTTACACCAACCACTTCAACAAATTTCCAGC 2269
 Qy 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTTPGlyTyr 80
 Db 2270 CAATCA---GGAGCTCGAACGACCAATCACTCTTTGGCTACAGCACCCCTTGGGGTAT 2326
 Qy 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTTPGlnArgLeuIleAen 100
 Db 2327 TTTGACTTCAACAGATTCCACTGCCACTTTTACCACGTCGTGCGCAAGACTCATCAAC 2386
 Qy 101 AsnAsnTTPGlyPheArgProLysArgLeuAsnPhelysLeuPheAsnIleGlnVallys 120
 Db 2387 AACAACTGGGATTCGGACCCACAGAGACTCAACTTCAAGCTCTTTAAATTCAGTCAAA 2446
 Qy 121 GluValThrThrAsnAspGlyValThrThrIleAlaAenAenLeuThrSerThrValGln 140
 Db 2447 GAGTTCAGCAGATGACGGTACCGACGAGATTGCCAATTAACCTTACCAGCAGGTTTCCAG 2506
 Qy 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
 Db 2507 GTGTTTACTGACTCGGAGTACAGCTCCGTCACGTCCTCGGCTCGGCGCATCAGGATGC 2566
 Qy 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAen 180
 Db 2567 CTCGCGCGTTCGCCAGCAGACGTCTTATGTGTCACAGATGAGATACCTTACCCTTGAC 2626
 Qy 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
 Db 2627 AACGGAGTTCAGGAGTAGGACGCTCTTCATTTTACTGCTGGAGTACTTTCTTCTCAG 2686
 Qy 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
 Db 2687 ATGCTGCGTACCGGAAACAACTTTACCTTCAGCTACACTTTTGGAGGAGCTTCTTTCAC 2746
 Qy 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAenProLeuIleAspGlnTyr 240
 Db 2747 AGCAGCTACGCTCACAGCAGAGTGGACCGTCTCATGAATCTCTCATGACCAAGTAC 2806
 Qy 241 LeuTyrTyrLeuAsnATgThrGlnAsnGlnSerGlySerAlaGlnAsnLeuAspLeuLeu 260
 Db 2807 CTGTATTACTTGGACGAGAAACAACTCCAGTGGGAACCAACCAAGCTCAAGGCTTCCAG 2866
 Qy 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTTPLeuProGlyPro 280
 Db 2867 TTTTCTCAGCGCGGAGCGAGTGACATTCGGGACCAAGCTCTAGGAATGCTCTCTCGAGCC 2926
 Qy 281 CysTyrArgGlnGlnAtqValSerIlysThrAspAenAenAenSerAsnAenPheThr 300
 Db 2927 TGTATCCCGCAGCAGCGAGTATCAAGACATCTCGCGGATAACAAACAGTGAATATCTCG 2986
 Qy 301 TTPThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAenProGlyThr 320
 Db 2987 TGGACTGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTGTTGAATCCGGGCCG 3046

QY 1 MetAlaSerGlyGlyGlyValAlaProMetAlaAspAsnAsnGluGlyValAlaAspGlyValGly 20
DB |||||
2809 ATGGCTACAGGCGAGTGGCGCCACCAATGCGAGACATACAGAGGGCGCGAGCGAGTGGGT 2868
QY 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 40
DB |||||
2869 AATTCCTCGGAAATTTGGCATTTGCCATTTCCACATGGATGGGCGACAGAGTTCATCACCACC 2928
QY 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrCysGlnIleSerSer 60
DB |||||
2929 AGCACCCGNACTGGGCGCTGGCCACCTACAAACACCACTCTCAAAACAATTTCCAGC 2988
QY 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
DB |||||
2989 CAATCA---GGAGCCTCGAAGACCAATCACTATTGGCTACAGCACCCCTTGGGGGTAT 3045
QY 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
DB |||||
3046 TTGTGACTTCAACAGATTCCACTGCCACTTTTCACCACGTCGATGTCGCAAGACTCATCAAC 3105
QY 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLeuPheAsnIleGlnValLys 120
DB |||||
3106 AACAACTGGGATTCGACCCCAAGAGACTCAACTTCAAGCTCTTTAAACATTCAGTCAAA 3165
QY 121 GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 140
DB |||||
3166 GAGGTTCACGACGAATGACGGTACGACGACGATTTGCCAATACTTACCGACAGGTTCCAG 3225
QY 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
DB |||||
3226 GTGTTTACTGACTCGGAGTACAGCTCCGCTAGCTCTCGGCTCGGCGCATCAAGGATGC 3285
QY 161 LeuProPhePheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
DB |||||
3286 CTCCCGCGGTTCCAGCAGACAGCTCTTCATGTGGCCACAGTATGGATACCTCACCTCGAAC 3345
QY 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
DB |||||
3346 AACGGGAGTCAGGAGTAGGAGCGCTCTTCATTTTACTGCTGGAGTACTTCTCTCTTCAG 3405
QY 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
DB |||||
3406 ATGCTGCGTACCGGAACAACATTTTACCTACCTACCTATTTTGGAGGAGCTTCTTTTCAC 3465
QY 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
DB |||||
3466 AGCAGCTAGCTTCACAGCAGAGTCTGGACCGTCTCATGATCTCTCATCGACAGTAC 3525
QY 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
DB |||||
3526 CTGTATTACTTGACGAGACAACACTCCAGTGGGAACCCACCACGAGTCAAGGCTTCAG 3585
QY 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
DB |||||
3586 TTTTCTCAGGCGCGAGCGAGTACACTCGGACCACTGAGGAGTCTGAGGAGTCTCTCTGAGCC 3645
QY 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
DB |||||
3646 TGTTACCGCCAGCAGGAGTATCAAGACATCTGCGGATACAAACACAGTGAATACTCG 3705
QY 301 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleAsnProGlyThr 320
DB |||||
3706 TGGACTGGAGTACCAAGTACCACTCAATGGCAGAGACTCTCTGTGTAATCCGGGCCG 3765
QY 321 AlaMetAlaSerHisLysAspGluAspLysPhePheProMetSerGlyValMetIle 340
DB |||||
3766 GCCATGGCAAGCCCAAGACCATGAAGAAAGTTTTCCTCAGAGCGGGTCTCATC 3825
QY 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
DB |||||
3826 TTTGGAGCAGGCTCAGAGAAACAAATGTGGACATTTGAAGAGTTCATGATTAAGAC 3885

QY 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
DB |||||
3886 GAAGAGGAATCAGGACACCAATCCCGTGGCTACGGAGCAGTATGTTCTGTATCTACC 3945
QY 381 AsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
DB |||||
3946 AACCTCCAGAGAGCAACACAGACACTACCGCAGATGTCAACACACAAAGGGCTTCTT 4005
QY 401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 420
DB |||||
4006 CCAGGCAATGGTCTGGCAGGACAGAGATGTACTCTTACGGGGCCCATCTGGGCAAGATT 4065
QY 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
DB |||||
4066 CCACACGAGCAGACATTTTACCCCTCTCCCTCATGGGTGATTCGGACTTAAACAC 4125
QY 441 ProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460
DB |||||
4126 CCTCCTCCACAGATTCTCATCAAGAACCCCGGTACTCGGAATCTTCGACCACCTTC 4185
QY 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
DB |||||
4186 AGTCGGCAAGTTTGTCTTCTTCATCACAGTACTCCACGGGACAGTCCAGGTGGAG 4245
QY 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrThr 500
DB |||||
4246 ATCGAGTGGAGCTGCAGAGGAAACAGCAACCGCTGGAATCCCGAAATTCAGTACACT 4305
QY 501 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 520
DB |||||
4306 TCCAACATCAACAAGTCTGTTAATGTGGACTTTACTGTGGACACTAATGGCGTGTATTC 4365
QY 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
DB |||||
4366 GAGCTCGCCCATTCGACACAGATACCTGACTGTAATCTG 4407

RESULT 10
AAF23750
ID AAF23750 standard; DNA; 4679 BP.
XX
AC AAF23750;
XX
DT 28-MAR-2001 (first entry)
XX
DE AAV2 DNA sequence.
XX
KW AAV2; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
KW atherosclerosis; sickle cell anaemia; thalassaemia;
KW blood clotting disorder; diabetes; ss.
XX
OS Adeno associated virus.
XX
PN US6156303-A.
XX
PD 05-DEC-2000.
XX
PF 11-JUN-1997; 97US-0873168.
XX
PR 11-JUN-1997; 97US-0873168.
XX
PA (UNIV) UNIV WASHINGTON.
XX
PI Russell DW, Rutledge EA;
XX
WPI; 2001-060164/07.
XX
PT Adeno-associated virus serotype 6 and viral vector derived from it for
PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
PT syndrome, sickle cell anemia, thalassaemia and diabetes -
XX
PS Claim 7; Fig 1; 50pp; English.
CC The present invention relates to adeno-associated virus serotypes. The

CC present sequence is the DNA sequence of one such serotype (AAV2). AAV2
 CC can be used to construct AAV viral vectors for use in gene therapy for a
 CC range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
 CC sickle cell anaemia, thalasassaemia, blood clotting disorders and diabetes.
 CC The AAV viral vectors have increased transduction efficiency of a
 CC particular host cell as the AAV virion containing the AAV vector genome
 CC can be modified to express a capsid protein of an AAV serotype that
 CC transduces the selected host cell.

XX Sequence 4679 BP; 1197 A; 1263 C; 1255 G; 964 T; 0 other;

Alignment Scores: 2,6e-227 Length: 4679
 Pred. No.: 2481.50 Matches: 445
 Score: 2481.50 Conservative: 37
 Percent Similarity: 90.26% Mismatches: 51
 Best Local Similarity: 83.33% Indels: 1
 Query Match: 85.39% Gaps: 1
 DB: 22

US-09-807-802A-17 (1-534) x AAF23750 (1-4679)

QY 1 MetAlaSerGlyGlyGlyAlaProMetAlaAspAsnGluGlyAlaAspGlyValGly 20
 DB 2809 ATGGCTACAGGCGAGTGGCGCACCAGTGGCAGCAATAGCGAGCGCGCGAGTGGGT 2868
 QY 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThr 40
 DB 2869 AATTCTCCGGAAATTTGGGATTCGGAATTCACATGGGATGGGCGACAGATCATACCCACC 2928
 QY 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrIlysGlnIleSer 60
 DB 2929 AGCACCCGAACTGGGCGCTGCCACCTACACCAACCACTCTACAAACAATTTCCAGC 2988
 QY 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
 DB 2989 CAATCA---GGAGCCTCGAAGCAATCACTATTGGCTACAGCACCCCTTTGGGGGTAT 3045
 QY 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
 DB 3046 TTTGACTTCAACAGATTCCACTGCCACTTTTACCACGCTGACTGGCAAGACTCATCAAC 3105
 QY 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 120
 DB 3106 AACCACTGGGGATTCGACCCACAGAGACTCACTTCAAGCTCTTTAACAATTCAGTCAAC 3165
 QY 121 GluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeuThrSerThrValGln 140
 DB 3166 GAGGTACAGCAGAAATGACGGTACGACGACGATTTGCCAATAACCTTACCAGCAGGTTACG 3225
 QY 141 ValPheSerAspSerGluTyrGlnLeuProTyrValIleGlySerAlaHisGlnGlyCys 160
 DB 3226 GTGTACTACCTCGGAGTACCAAGCTCCCGTACCTCTCGGCTCGGCGCATCAAGATGC 3285
 QY 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
 DB 3286 CTCGCGCGTTCACAGCAGAGCTTTCATGTGTGCCAGATGATGATACCTCACCTCGAAC 3345
 QY 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
 DB 3346 AACGGGAGTCAGGCGAGTGGAGCGCTCTTCAATTTTACTGCGCTGGAGTACTTTCCTCTCAG 3405
 QY 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
 DB 3406 ATGCTGGTACCGGAAACAATTTTACCTTCAGCTACACTTTTGGAGCGTTCCTTCCAC 3465
 QY 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
 DB 3466 AGCAGCTACGCTCACAGCCAGAGTCTCGACCGTCTCATGAATCTCTCATCGCACCAGTAC 3525
 QY 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
 DB 3526 CTGTATTACTTGACGAGCAACAAACATCCCAAGTGGAAACCCAGCAGCTCAAGGCTTCAG 3585

QY 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
 DB 3586 TTTTCTCAGGCGGAGCGAGTGCATTCGGGACCAAGTCTAGAACTGGCTTCTGGGACCC 3645
 QY 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
 DB 3646 TGTACCAGCCAGCAGCGAGTATCAAGACATCTCGGATAACAACAACAGTGAATACTCG 3705
 QY 301 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr 320
 DB 3706 TGGACTGGAGCTACCAAGTACCCTCAATGGCAGAGACTCTCTGGTGAATCCGGGCCCG 3765
 QY 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
 DB 3766 GCCATGGACCCACAGGAGCGATGAGAAAGTTTTTCTTCAGAGCGGGTTCATC 3825
 QY 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
 DB 3826 TTTGGGAAGCAAGGCTCAGAGAAAACAATGTGCACATTTGAAAGGTCATGATTACAGAC 3885
 QY 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
 DB 3886 GAAGAGGAAATCAGGACAAACAATCCCGTGGCTACGGAGCAGATGTTCTGTATCTACC 3945
 QY 381 AsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
 DB 3946 AACCTCCAGAGGCAACAGACAGCTACCCAGATGTCACACACAAGGCGTCTT 4005
 QY 401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 420
 DB 4006 CCAGCATGCTGTGCAGACAGAGATGTGTACCTTTCAGGGGCCCATCTGGGCAAGATT 4065
 QY 421 ProHisThrAspGlyHisPheHisPheSerProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
 DB 4066 CCACACGGGACGGACATTTTCCACCCCTCTCCCTCATGGGTGGATTCCGACTTAACAC 4125
 QY 441 ProProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460
 DB 4126 CCTCTCCACAGATTTCTCATCAAGAACACCCCGGTACCTCGGATCTTCGACCACTTC 4185
 QY 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
 DB 4186 AGTCGGCAAAAGTTTGTCTTCTCATCACAGTACTCCACGGGACAGGTTCAGCGTGAG 4245
 QY 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrThr 500
 DB 4246 ATCGAGTGGAGCTCGAAGGAAACAGAACCTGGAAATCCCGAAATTCAGTACACT 4305
 QY 501 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 520
 DB 4306 TCCAACTACAAAGTCTGTTAATGTGGACTTTTACTGTGGACACTAATGGCGGTATTCA 4365
 QY 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
 DB 4365 GAGCCTCGCCCATTTGGCACCAGATACCTGACTCGTAATCTG 4407

RESULT 11

ABK89694
 ID ABK89694 standard; DNA; 4679 BP.
 XX AC ABK89694;
 XX DT 05-NOV-2002 (first entry)
 XX DE Adeno-associated virus 2 (AAV2) vector.
 XX Adeno-associated virus 2 vector; AAV2; ds; cyclic; circular; cancer;
 KW VP1 capsid; heparin-sulphate proteoglycan; vaccine; immune response;
 KW ovarian cancer.
 XX Adeno-associated virus 2.
 OS
 XX FH Location/Qualifiers

CDS 2203..4410
 /*tag= a
 /product= "Adeno-associated virus 2, VP1 capsid protein"
 CDS 2614..4410
 /*tag= b
 /product= "Adeno-associated virus 2, VP2 capsid protein"
 /transl_except= (pos:2614..2616, aa:Met)
 2809..4410
 /*tag= c
 /product= "Adeno-associated virus 2, VP3 capsid protein"
 WO200253703-A2.
 11-JUL-2002.
 04-JAN-2002; 2002WO-US00152.
 05-JAN-2001; 2001US-260124P.
 (CHIL-) CHILDRENS HOSPITAL INC.
 Bartlett JS;
 WPI; 2002-583608/62.
 P-PSDB; AAU98974, AAU98975, AAU98976.
 New adeno-associated virus vector comprises a biotinylated capsid or capsid protein with an amino acid insertion in the VP1 capsid, useful as a vaccine or for transferring a therapeutic peptide to a cancer cell
 Disclosure; Page 46-48; 57pp; English.
 The invention relates to an adeno-associated virus (AAV) vector (I) comprising a biotinylated capsid or capsid protein (II) with an amino acid insertion following the capsid amino acid at position 139, 161, 588 or 657 in the VP1 capsid. The AAV vector comprises a capsid protein containing one or more amino acid insertions that ablate the ability of the vector to bind heparin-sulphate proteoglycan and allow the vector to use a cellular receptor not used by wild type AAV. Modified (I) are useful as vaccines to elicit immune responses to amino acids, where the response can be protective and/or therapeutic. (I) may be used to transfer a therapeutic peptide to a cancer cell, particularly to an ovarian cancer cell. The present sequence represents the adeno-associated virus 2 (AAV2) vector used in to make modified AAV2 vectors.
 SQ Sequence 4679 BP; 1198 A; 1262 C; 1255 G; 964 T; 0 other;
 Alignment Scores:
 Pred. No.: 2, 6e-227 Length: 4679
 Score: 2481.50 Matches: 445
 Percent Similarity: 90.26% Conservative: 37
 Best Local Similarity: 83.33% Mismatches: 51
 Query Match: 85.39% Indels: 1
 DB: 24 Gaps: 1
 US-09-807-802A-17 (1-534) x ABK89694 (1-4679)
 QY 1 MetAlaSerGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGly 20
 DB 2809 ATGGCTACAGGCGAGTGGCGCAATGCGAGCAATAACGAGGGCGCGAGTGGGT 2868
 QY 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 40
 DB 2869 AATTCCTCGGAAATTTGGCATTTGCCATTTCCACATGGATGGGCGCAGAGTTCATCACC 2928
 QY 41 SerThrArgTrpAlaLeuProThrTyrAsnAsnHisLeuTyrCysGlnIleSerSer 60
 DB 2929 AGCACCCGAACTTGGGCGCTGGCCACCTTACAAACACCACTCTACAAACAAATTTCCAGC 2988

QY 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
 DB 2989 CAATCA--GGAGCCTCGAACGACAATCACTACTTTGGCTACAGCACCCCTTGGGGGTAT 3045
 QY 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
 DB 3046 TTTGACTTCAACAGATTCCCACTGCGCACTTTTCCACCACTGACTGGCAAAAGACTCATCAAC 3105
 QY 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 120
 DB 3106 AACAACTGGGGATTCCGACCCAGAGACTCACTTCAAGCTCTTTTACATTCAGTCAAA 3165
 QY 121 GluValThrThrAsnAspGlyValThrIleAlaAsnLeuThrSerThrValGln 140
 DB 3166 GAGGTCAACGAGATGACGGTACGACGATTCGAATTAACCTTACGACACGGTTCAG 3225
 QY 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
 DB 3226 GTGTTTACTGACTCGGAGTACCAGCTCCGCTACGCTCTCGGCTCGGCGCATCAAGGATGC 3285
 QY 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
 DB 3286 CTCCCGCGGTTCCAGCAGAGCTCTTCATGTGTGCGACAGTATGATACCTCACCCTGAAAC 3345
 QY 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
 DB 3346 AACGGAGTCAGGCAGTAGGACGCTCTTCAATTTTACTGCTGGAGTACTTTCTCTCTCAG 3405
 QY 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
 DB 3406 ATGCTGGGTACCGGAAACAACTTTTACCTTACGCTACACTTTTGGAGGAGTCTCTTTCCAC 3465
 QY 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
 DB 3466 AGCAGCTACGCTACAGCCAGAGCTCGGACGCTCTCATGAATCTCTCATCGACCATGATAC 3525
 QY 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
 DB 3526 CTGTATTACTTGGACAGAACAAACAACTCCAAGTGGAAACCAACGACGATCAAGGCTTCAG 3585
 QY 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
 DB 3586 TTTTCTCAGGCGGAGGAGTGCATTTCCGGACCACTGAGAACTGCTTCTCTGGACCC 3645
 QY 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
 DB 3646 TGTATCCGCCAGCAGCAGTATCAAGACATCTCGGATTAACAACACAGTGAATACTCG 3705
 QY 301 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr 320
 DB 3706 TGGACTGGAGCTACCAAGTATACCTCAATGGCAGAGACTCTCTGTTGAATCCGGGCCG 3765
 QY 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
 DB 3766 GCCATGGCAAGCCCAAGGACGATGAAGAAAGTTTTCCTCAGAGCGGGTCTCTCATC 3825
 QY 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
 DB 3826 TTTGGGAAGCAAGGCTCAGAGAAAACAATGTGCACATTTGAAGAGTTCATGATACAGAC 3885
 QY 361 GluGluLysLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
 DB 3886 GAAGAGAAATCAGGACAAACCAATCCGCTGCTACGGAGAGTATGGTCTCTGTATCTACC 3945
 QY 381 AsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
 DB 3946 AACCTCAGAGAGCAACAGACAGAGTACCAGAGATGTCAACACAAAGGCGTCTCT 4005
 QY 401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 420
 DB 4006 CCAGGCTGCTGTCGAGGACAGAGATGTGTACCTTCCAGGGGCCCATCTCGGCAAGATT 4065
 QY 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440


```

Db      3646  TGTATCCGCGCAGCGAGTATCAAGACATCTGGCGGATAACAACAACAGTGAATACCTCG 3705
QY      301  TTPThrGlyAlaSerIySvrrAenLeuAenGlyArqGluSerIlelleAenProGlyThr 320
Db      3706  TGGACTGGAGCTACCAAGTACCACTCAATGGCAGAGACTCTGTGTGAATCCGGGCCG 3765
QY      321  AlaMetAlaSerHiglyAspAspGluAspIysPhePheProMetSerGlyValMetIle 340
Db      3766  GCATGGCAAGCCACCAAGCAGCATGAAGAAAGTTTTTCTCTCAGAGCGGGTCTCATC 3825
QY      341  PheGlyIysGluSerAlaGlyAlaSerAenThrAlaLeuAspAenValMetIleThrAsp 360
Db      3826  TTTGGGAACCAAGGCTCAGAGAAACAATGTGGACATTGNAAGAGTCTATGATTACAGAC 3885
QY      361  GluGluGluIleIysAlaThrAenProValAlaThrGluArgPheGlyThrValAlaVal 380
Db      3886  GAAGAGGAATCAGGCACACCAATCCCGTGGCTACGGAGCATATGTTCTGTATCTACC 3945
QY      381  AenPheGlnSerSerThrAspProAlaThrGlyAspValHigAlaMetGlyAlaLeu 400
Db      3946  AACCTCCAGAGAGGCAACACAGCAGCTACCGCAGATGTCAACACACAAGCGCTTCTT 4005
QY      401  ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleThrAlaIysIle 420
Db      4006  CCAGGCGATGGTGGCAGACAGAGATGTGTACTTTCAGGGGCCCATCTGGGCAAGATT 4065
QY      421  ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuIysAen 440
Db      4066  CCACACACGAGCGGACATTTTACCCCTCTCCCTCATGGTGGATTCGACATTAAACAC 4125
QY      441  ProProProGlnIleLeuIleIysAenThrProValProAlaAenProProAlaGluPhe 460
Db      4126  CCTCTCCACAGATTCTCATCAAGAACACCCCGGTACCTGCGAATCCTTCGACACCTTC 4185
QY      461  SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
Db      4186  AGTGGCGGAAGTTTGCTTCTTCATCACACAGTACTCCACGGGACAGTCCAGCGTGGAG 4245
QY      481  IleGluTrpGluLeuGlnLysGluAenSerLysArqTrpAenProGluValGlnTyrThr 500
Db      4246  ATCGATGGGAGCTGCAGAGGAAACACAGCAACCGCTGAATCCCGAATTCAGTACACT 4305
QY      501  SerAsnTyrAlaIysSerAlaAenValAspPheThrValAspAenAenGlyLeuTyrThr 520
Db      4306  TCCAACTACAACAAGTCTGTTAATGTGGACTTTACTGTGGACACTAATGGCGTGTATTCA 4365
QY      521  GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
Db      4366  GAGCCTCGCCCATTTGGACCAACAGATACCTGACTCGTGAATCTG 4407

```

RESULT 13

AAH26326

ID AAH26326 standard; DNA; 7557 BP.

XX AAH26326;

XX 02-OCT-2001 (first entry)

DT Adeno-associated virus packaging plasmid pfloxAAV.

DE AAV; vector; pfloxAAV2; packaging cell line; gene therapy; ds.

XX Adeno associated virus.

XX WO200155361-A2.

XX 02-AUG-2001.

XX 26-JAN-2001; 2001WO-US02709.

XX 26-JAN-2000; 2000US-0178536.

XX

(CHIR) CHIRON CORP.

Hardy SF;

WPI; 2001-483239/52.

Producing recombinant adeno-associated virus (rAAV) vector, by stably

infecting eukaryotic host cell with rAAV vectors, helper adenovirus or

herpes virus, recombinant adenovirus or herpes vectors -

Example 1; Page 54-57; 63pp; English.

The present sequence is that of adeno-associated virus (AAV) packaging plasmid pfloxAAV2. The following AAV packaging plasmids were prepared: pK5repCap, a non-replicating control plasmid containing rep and cap genes but no inverted terminal repeats (ITRs); pAV2 (see AAH26325), containing a wild-type AAV genome; and pfloxAAV, a version of pAV2 with 2 loxP sites inserted such they flank the rep and cap genes and thus separate the ITRs from these genes. Each was combined with pCMV GFP (see AAH26324), a plasmid with a green fluorescent protein expressing recombinant AAV (rAAV) genome, and used to transfect 293 or Cre8 cells. After 6 hr, the medium was changed and wild-type adenovirus type 5 was added. Virus particles were harvested 3 days later. Plasmid pAV2 produced mostly AAV and a low yield of GFP vector with no effect of Cre recombinase. pfloxAAV packaged as much GFP vector as the control plasmid and while the amount of GFP vector was not affected by Cre recombinase, the amount of floxAAV in the product was reduced to 1/10 of the 293 value by the action of Cre recombinase. In 293 cells it was evident that the GFP vector did not compete with wild-type AAV for replication or packaging, while the GFP vector was able to complete against floxAAV. Methods and compositions are provided for producing rAAV vector particles by: (a) introducing into a host cell (i) pfloxAAV, (ii) a recombinant viral vector encoding plasmid, and (iii) a plasmid encoding herpesvirus, cytomegalovirus or adenoviral functions, or a herpesvirus, cytomegalovirus or adenovirus itself, to produce a flox AAV particles and rAAV particles; and (b) introducing into a second host cell (i) the rAAV particles or (a), (ii) a vector that directs expression of Cre, and (ii) a vector which directs expression of herpes virus, cytomegalovirus or adenovirus helper functions, such that rAAV vector particles are produced. The vectors are useful for in vivo or in vitro gene therapy and also for in vitro recombinant protein production.

Sequence 7557 BP; 1970 A; 1940 C; 1953 G; 1694 T; 0 other;

Alignment Scores:

Pred. No.: 5,16e-227 Length: 7557

Score: 2481.50 Matches: 445

Percent Similarity: 90.26% Conservative: 37

Best Local Similarity: 83.33% Mismatches: 51

Query Match: 85.39% Indels: 1

DB: 22 Gaps: 1

US-09-807-802A-17 (1-534) x AAH26326 (1-7557)

QY 1 MetAlaSerGlyGlyGlyAlaProMetAlaAspAenAenGlyAlaAspGlyValGly 20
 Db 2837 ATGGCTACAGCAGTGGCGCACCACCAATGCACACAATAACGAGGGCGCGAGTGGGT 2896

QY 21 AenAlaSerGlyAenTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 40
 Db 2897 AATTCTCGGAAATTTGGCATTGGCATTTCCACATGGATGGGCGCAGAGTGCATCCACC 2956

QY 41 SerThrArgThrTrpAlaLeuProThrTyrAenAenHisLeuTyrLysGlnIleSerSer 60
 Db 2957 AGCACCCGAACCTGGGCCCTCCACCTTACCAACCAACACCTCTTCAACAACAATTTCCAGC 3016

QY 61 AlaSerThrGlyAlaSerAsnAspAenHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
 Db 3017 CAATCA---GGAGCCTCGAACGACAATCCTACTTTGGCTACAGCACCCCTTGGGGGTAT 3073

QY 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspProGlnArgLeuIleAsn 100
 DB 3074 TTTGACTTCAACAGATTCCACTGCCACTTTTACCACGAGCTGGCAAGACTCATCAAC 3133
 QY 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 120
 DB 3134 AACAACTGGGGATTCCGACCCCAAGAGACTCAACTTCAAGCTCTTTAACTTCAAGTCAA 3193
 QY 121 GluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeuThrSerThrValGln 140
 DB 3194 GAGGTCAAGCAGAAATGACGGTACGACGACGAGTTCGCAATTAACCTTACCAGCAGCGTTCC 3253
 QY 141 ValPheSerAspSerGluTyrGlnLeuProTyrValIleGlySerAlaHisGlnGlyCys 160
 DB 3254 GTGTTTACTGACTCGGAGTACGAGCTCCCTAGCTCTCGGCTCGGCGCATCAGGATGC 3313
 QY 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
 DB 3314 CTCCCGCGTTCCAGCAGACGCTTCTCATGTCGCCACAGTATGATACCTCACCCCTGAAC 3373
 QY 181 AsnGlySerGlnAlaValGlyArgSerPheTyrCysLeuGluTyrPheProSerGln 200
 DB 3374 AACGGAGTCAGGACGATGAGACGCTCTTCAATTTTACTGCTGGAGTACTTTCCTCTCAC 3433
 QY 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
 DB 3434 ATGCTGGTACCAGAAACAATTTTACTTTCAGCTACATTTTTCAGGACGTTCTTTCAC 3493
 QY 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
 DB 3494 AGCAGCTACGCTCACAGCCAGAGTCTGGACCGTCTCATGAATCTCTCATCGACCAAGTAC 3553
 QY 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeu 260
 DB 3554 CTGTATTACTTGTAGCAGAAACAACATCTCAAGTGGAAACACCGCAGTCAAGGCTTCAG 3613
 QY 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
 DB 3614 TTTTCTCAGCGCGGAGTGACATTCGGGACCACTCTAGGAAGTGGCTTCTTGGACCC 3673
 QY 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
 DB 3674 TGTATTCCCGCCAGCGAGTATCAAGACATCTCGCGATACCAACAACAGTGAATATCTCG 3733
 QY 301 TrpThrGlyValSerLysTyrAsnLeuAsnGlyArgGluSerIleLeuLeuAsnProGlyThr 320
 DB 3734 TGGACTGGAGCTACCAAGTACCCTCAATGGCAGAGACTCTCTGGTGAATCCGGGCCCG 3793
 QY 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
 DB 3794 GCCATGCCAAGCCACAGGACGATGAAGAAAGTTTTCCTCAGAGCGGGTCTCATC 3853
 QY 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
 DB 3854 TTTGGGAAGCAAGGCTCAGAGAAACAATGTGGACATTTGAAAGGTCTATGATTACAGAC 3913
 QY 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
 DB 3914 GAAGAGAAATCAGGACCAACAATCCCGTGGCTACGCGAGTATGTTCTGTATATACC 3973
 QY 381 AsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
 DB 3974 AACCTCCAGAGAGGCAACAGACAGCAGCTACCGCAGATGTCAACACACAAGCGTCTT 4033
 QY 401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 420
 DB 4034 CCAAGCATGTCTGGCAGGACAGAGATGTGTACTTTCAGGGGCCCATCTCTGGCAAGAT 4093
 QY 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
 DB 4094 CCACACAGGACGACATTTTCAACCTCTCCCTCATGGGTGGATCGGACTTAACAC 4153
 QY 441 ProProProGlnIleLeuIleLysAsnThrProValProAlaAsnProAlaGluPhe 460

DB 4154 CTTCTCTCCACAGATTCTCATCAGAACACCCCGGTACCTGGAATCTTTCGACCACTTC 4213
 QY 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
 DB 4214 AGTGGCGCAAGTTTGTCTTCTTCTATCACAGTACTCCACGGGACAGGTTCAGCGTGGAG 4273
 QY 481 IleGluTyrGluLeuGlnLysGluAsnSerLysArgTyrAsnProGluValGlnTyrThr 500
 DB 4274 ATCGAGTGGAGCTCCAGAGAAACAGCAACCGCTGGATCCCGAATTCAGTACACT 4333
 QY 501 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 520
 DB 4334 TCCAACTACAACAAGTCTGTTAATGTGGACTTTTACTGTGGACACTAATGGCGGTATTCA 4393
 QY 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
 DB 4394 GAGCCTCGCCCAATTTGGCACCAGATACCTGACTCGTAATCTG 4435
 RESULT 14
 AAH26325
 ID AAH26325 standard; DNA; 8698 BP.
 XX
 AC AAH26325;
 XX
 DT 02-OCT-2001 (first entry)
 XX
 DE Wild-type adeno-associated virus genome plasmid pAV2.
 XX
 KW AAV; vector; pAV2; packaging cell line; gene therapy; ds.
 XX
 OS Adeno associated virus.
 XX
 PN WO200155361-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 26-JAN-2001; 2001WO-US02709.
 XX
 PR 26-JAN-2000; 2000US-0178536.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI Hardy SF;
 XX
 DR WPI; 2001-483239/52.
 XX
 PT Producing recombinant adeno-associated virus (rAAV) vector, by stably
 infecting eukaryotic host cell with rAAV vectors, helper adenovirus or
 herpes virus, recombinant adenovirus or herpes vectors -
 XX
 PS Example 1; Page 51-54; 63pp; English.
 XX
 CC The present sequence is that of adeno-associated virus (AAV)
 packaging plasmid pAV2, which contains a wild-type AAV type 2
 genome. The following AAV packaging plasmids were prepared:
 CC pK5repCap, a non-replicating control plasmid containing rep and
 cap genes, but no inverted terminal repeats (ITRs); pAV2; and
 CC pflxAAV (see AAH26326), a version of pAV2 in which 2 inserted loxP
 sites flank the rep and cap genes and thus separate the ITRs from
 CC these genes. Each was combined with pCMV GFP (see AAH26324), a
 CC plasmid with a green fluorescent protein expressing recombinant AAV
 (rAAV) genome, and used to transfect 293 or Cre8 cells. After 6
 CC hr, the medium was changed and wild-type adenovirus type 5 was
 CC added. Virus particles were harvested 3 days later. Plasmid pAV2
 CC produced mostly AAV and a low yield of GFP vector with no effect
 CC of Cre recombinase. pflxAAV packaged as much GFP vector as the
 CC control plasmid and while the amount of GFP vector was not
 CC affected by Cre recombinase, the amount of floxAAY in the product
 CC was reduced to 1/10 of the 293 value by the action of Cre
 CC recombinase. In 293 cells it was evident that the GFP vector did
 CC not compete with wild-type AAV for replication or packaging, while
 CC the GFP vector was able to compete against floxAAY. Methods

CC and compositions are provided for producing RAAV vector particles
 CC by: (a) introducing into a host cell (i) p10xAAV, (ii) a
 CC recombinant viral vector encoding plasmid, and (iii) a plasmid
 CC encoding herpesvirus, cytomegalovirus or adenoviral functions, or
 CC a herpesvirus, cytomegalovirus or adenovirus itself, to produce
 CC a RAAV particle; and (b) introducing into a
 CC second host cell (i) the RAAV particles or (a), (ii) a vector
 CC that directs expression of Cre, and (iii) a vector which directs
 CC expression of herpes virus, cytomegalovirus or adenovirus helper
 CC functions, such that RAAV vector particles are produced. The
 CC vectors are useful for in vivo or in vitro gene therapy and also
 CC for in vitro recombinant protein production.

XX Sequence 8698 BP; 2117 A; 2368 C; 2306 G; 1907 T; 0 other;

Alignment Scores:

Pred. No.: 6 31e-227 Length: 8698
 Score: 2481.50 Matches: 445
 Percent Similarity: 90.26% Conservative: 37
 Best Local Similarity: 83.33% Mismatches: 51
 Query Match: 85.39% Indels: 1
 DB: 22 Gaps: 1

US-09-807-802A-17 (1-534) x AAH26325 (1-8698)

QY 1 MetaSerGlyGlyGlyAlaProMetAlaAspAsnGluGlyAlaAspGlyValGly 20
 DB ATGGCTACAGGAGTGGCGACCAATGGCAGACAATAACAGGGCGCGAGTGGGT 2868
 QY 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThr 40
 DB AATTCCTCGGAAATTTGGCAATTTGGCATTTCCATGATGGCGAGAGTCAACACC 2928
 QY 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSer 60
 DB AGCACCAGAACTTGGGCTTGGCCCTGACCTACACACACCCTCTACAAACAAATTTCCAGC 2988
 QY 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
 DB CAATCA--GGAGCCTCGAAGCACAATCACTACTTTGGCTACAGCACCCCTTGGGGGTAT 3045
 QY 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
 DB TTGACTTCAACAGATTCCACTGCCACTTTTCCACACGTCGAGTGGCAAGACTCAATCAAC 3105
 QY 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnAsnPheLysLeuPheAsnIleGlnValLys 120
 DB AACAACTGGGGATTCGACCCCAAGAGACTCACTTCAAGCTCTTTAAACATTCAAGTCAAA 3165
 QY 121 GluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeuThrSerThrValGln 140
 DB GAGGTCACGCAGAAATGACGGTACGACGAGCGATTTGCCAATAAATTAACAGCAGCGTTTCA 3225
 QY 141 ValPheSerAspSerGlyTrpGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
 DB GTGTTTACTGACTCGGAGTACAGCTCCCGTACGCTCTCGCTCGCGCGCATCAAGGATGC 3285
 QY 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
 DB CTCCCGCGGTTCCACAGACAGCTTTCATGGTGGCCACAGTATGGATACCTTCACTTCAAC 3345
 QY 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
 DB AACGGGAGTCAGGACAGTGGAGCTTTCATTTTACTGCTGGAGTACTTCTCTTCTCAG 3405
 QY 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
 DB ATGCTCGTACCGGAAACAACTTATCTTACCTACAGTACACTTTTGGAGAGCGTCTCTTCCAC 3465
 QY 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
 DB AGCAGCTACGCTCACAGCCAGAGTCTGGACCGCTCTCATGATCTCTCATCCACAGTAC 3525

QY 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
 DB CTGTATTACTTGACGACAAACAACTCAAGTGGAAACACACACAGCTCAAGGCTTCAG 3585
 QY 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
 DB TTTTCTCAGGCGGAGCGAGTGCATTTGGGACCAAGTCTAGGAATGCTGCTTCTGGACCC 3645
 QY 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
 DB TGTACCGCCAGCAGCGAGTATCAAGACATCTCGGATACACACACAGTGAATACTCG 3705
 QY 301 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr 320
 DB TGGACTGGAGCTACCAAGTATCACTCAATGGCAGAGACTCTCTGGTGAATTCGGGCGCCG 3765
 QY 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
 DB GCCATGGCAAGCCCAAGGACGATGAAGAAAGTTTTTCTCAGAGCGGGTTCATC 3825
 QY 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
 DB TTTGGGAAGCAAGCTCAGAGAAAACAATGTGCACATTTGAAAAGTTCATGATTACAGAC 3885
 QY 361 GluGluGluLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
 DB GAAGAGGAAATCAGCAACAACCAATCCGTGGTACGGAGAGTATGGTTCGTATATACC 3945
 QY 381 AsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
 DB AACCTCCAGAGAGCAACAGCAAGCAGCTACCCAGATGTCAACACACAAAGGCGTCTT 4005
 QY 401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 420
 DB CCAGCATGGTCTCGCAGGACAGAGATGTGTACCTTCAGGGGCCCATCTGGGCAAGATT 4065
 QY 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
 DB CCACACAGCGGACGACATTTTCCACCTCTCCCTCATGGGTGGATTCGGACTTAAACAC 4125
 QY 441 ProProGlnIleLeuLysAsnThrProValProAlaAsnProAlaGluPhe 460
 DB CCTCTCCACAGATCTCATCAAGACACCCCGGTACTCTGCGAATCTCTCCAGCACCTTC 4185
 QY 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
 DB AGTGGCGCAAGTTTGTCTTCTTCATCACAGTACTCCACGGGACAGGTGAGCGTGAG 4245
 QY 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrThr 500
 DB ATCAGTGGGAGCTGCAGAGGAAACAGCAACGCTGGAATCCCGAATTCAGTACT 4305
 QY 501 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 520
 DB TCCAACTACACAGTCTGTTAATGTGGACTTACTGTGGACACTTAATGGGCGGTATTCA 4365
 QY 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
 DB GAGCTCGCCCAATTTGGCACCAGATACCTGACTCGTAATCTG 4407

RESULT 15

AAAT09008

ID AAT09008 standard; DNA; 4680 BP.

XX AC AAT09008;

XX XX

DT 05-JUL-1996 (first entry)

DE Wild-type adeno-associated virus 2 genome.

XX adeno-associated virus 2; AAV2; inverted terminal repeat; viral vector;

KW DNA delivery; immunodeficiency virus protein; immunity; human; simian;

KW neurological disorder; therapy; ss.

XX OS Adeno-associated virus.
 XX PN W09534670-A2.
 XX PD 21-DEC-1995.
 XX XX 06-JUN-1995; 95WO-US07178.
 XX PF 06-JUN-1994; 94US-0254358.
 XX PR (CHIL-) CHILDRENS HOSPITAL INC.
 XX PA Johnson PR;
 XX PI WPI; 1996-049697/05.
 XX DR Recombinant adeno-associated virus genome contg. protein encoding
 XX XX DNA - flanked by inverted terminal repeats, for use in vaccines or
 XX PT for treatment of neuro-degenerative disease
 XX PS Example 1; Page 27-29; 44pp; English.
 XX CC The present sequence is that of the wild-type adeno-associated virus
 CC (AAV) 2, a replication-deficient parvovirus. Cis-acting sequences
 CC directing viral DNA replication (ori), encapsidation/packaging (pk) and
 CC host cell chromosome integration (int) are contained within the ITRs
 CC (inverted terminal repeats of 145 nucleotides). When AAV infects a
 CC human cell, the viral genome integrates into chromosome 19 resulting in
 CC latent infection of the cell. Prodn. of infectious virus does not occur
 CC unless the cell is infected with a helper virus (e.g., adenovirus or
 CC herpes-virus). AAV possesses unique features that make it attractive as
 CC a vector for delivering foreign DNA to cells. A vector including a
 CC recombinant AAV genome contg. a simian immunodeficiency virus (SIV) rev
 CC and envelop (gpl60) gene cassette was constructed from an existing plasmid
 CC designated psb201. The plasmid contains a modified wild-type AAV2
 CC genome. Specifically, two XbaI sites were added via linker addition at
 CC sequence positions 190 and 484. These sites are internal to 191 bp ITRs
 CC which include the 145 bp ITRs of the AAV genome. The insertion of these
 CC sites allows the complete removal of the internal 4.3 kb fragment contg.
 CC the AAV rep-cap genes upon XbaI digestion of the plasmid. (See also
 CC AAT09009-10).
 XX SQ Sequence 4680 BP; 1198 A; 1265 C; 1254 G; 963 T; 0 other;

Alignment Scores:
 Pred. No.: 2,78e-223 Length: 4680
 Score: 2439.50 Matches: 441
 Percent Similarity: 89.16% Conservative: 36
 Best Local Similarity: 82.43% Mismatches: 56
 Query Match: 83.95% Indels: 2
 DB: 17 Gaps: 1

US-09-807-802A-17 (1-534) x AAT09008 (1-4680)

Qy 1 MetAlaSerGlyGlyAlaProMetAlaAspAsnGluGlyAlaAspGlyValGly 20
 Db 2809 ATGGCTACAGGCAGTGGCGGACCAATGGCAGCAATACGAGGCGCGGAGTGGGT 2868
 Qy 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 40
 Db 2869 AATTCTCCCGAATTTGGCATTTGGATTCCATCGATGGATGGGCGACAGATCATCCACC 2928
 Qy 41 SerThrArgThrTrpAlaLeuProThrTrpAsnAsnHisLeuTyrLysGlnIleSerSer 60
 Db 2929 AGCACCGGAACTGGGCGCTCCACCTACCAACACCAACCTCTACAAACAAATTTCCAGC 2988
 Qy 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
 Db 2989 CAATCA--GGAGCCTCGAACGACAAATCACTTTTGCTTACACACCCCTTGGGGGTAT 3045
 Qy 81 PheAspPheAsnArgPheHisCysPheSerProArgAspTrpGlnArgLeuIleAsn 100

Db 3046 TTTGACTTCAACAGATTCCACTGCCACTTTTCCACCAGTGCAGTGCACAAAGACTCATCAAC 3105
 Qy 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 120
 Db 3106 AACAACTGGGATTTCCGACCCAGAGACTCAACTTCAAGCTCTTTAAACATTCAAGTCAAA 3165
 Qy 121 GluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeuThrSerThrValGln 140
 Db 3166 GAGGTACGACAGATGAGGTACGACGACGATTTGCCAATACCTTACCAGCAGCGTTCCAG 3225
 Qy 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
 Db 3226 GTGTTTACTGACTCGGAGTACAGCTCCGCTACGCTCTCGCTCGGCGCATCAGGATGC 3285
 Qy 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
 Db 3286 TCCCGCGGTTTCCAGCAGACGCTTTCATGTGTCACAGTATGGATATCTTCCCTTGAC 3345
 Qy 181 AsnGlySerGlnAlaValGlyArgSerPheTyrCysLeuGluTyrPheProSerGln 200
 Db 3346 AACGGAGTACGACGATGAGGAGCTCTTCATTTTACTGCTGGAGTACTTTCTCTCTCAG 3405
 Qy 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
 Db 3406 ATGCTGCGTACCGGAAACAACTTTTACCTTACCTTACCTTTTGGAGGAGCTTCTTCCAC 3465
 Qy 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
 Db 3466 AGCAGTACGCTACAGCAGAGTCTGACCGCTCTCATGAATCTCTCATCAGCAGTAC 3525
 Qy 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
 Db 3526 CTGTATTACTTGAGCAGACAAACACTCCAAGTGGACACCAACAGCTCAAGGCTTCCAG 3585
 Qy 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
 Db 3586 TTTTCTCGGCGGACGAGTGCATCTCGGACCACTCTAGGAACCTGCTCTCTCGGACCC 3645
 Qy 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
 Db 3646 TGTTACCAGCAGCAGGATTCAAAGACATCTCGGATACACACACACGATGAATACTCG 3705
 Qy 301 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGlySerIleLeuAsnProGlyThr 320
 Db 3706 TGGACTGGAGTACCACTCACTCAATGSCAGAGACTCTCTGTGTAATCCGGGCGCC 3765
 Qy 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
 Db 3766 GCCATGGCAAGCCACAAAGACGATGAAGAAAAAGTTTTTCTCAGAGCGGGTTCTCATC 3825
 Qy 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
 Db 3826 TTTGGNACGAGGCTCAGAGAAACAAATGTGAACATTTGAAGGTCTATGATTACAGAC 3885
 Qy 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
 Db 3886 GAAGAGAAATCGGAACAACTCCGTTGGCTACGAGCAGTATGTTTCTGTTATCTACC 3945
 Qy 381 AsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
 Db 3946 AACCTCCAGAGAGCCACAGACAGCAGCTACCGCAGATGTCAACACACAAAGCGTTCT 4005
 Qy 401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 420
 Db 4006 CAGGCTGTGTCTGGCAGCAGAGATGTGATCTTTCAGGGGCCCATCTGGGCAAGATT 4065
 Qy 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
 Db 4066 CCACACAGCAGCAGCATTTTCACTCTCCCTCATGGTGGATTCGAGCTTAAACAC 4125
 Qy 441 ProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460
 Db 4126 CCTCTCCAGATTCTCATCAAGAACACCCCGGTACCTGCGAATCTTCTCGACCACTTC 4185

```
QY 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGln-ValSerValGI 480
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4186 AGTGGCGCAAGCTTGTCTTCATCACAGTACTCCACGGGACACGGTCAGCGTGA 4245
QY 480 uIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrTh 500
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4246 GATCGAGTGGGAGCTGCAGAGGAAACACAGCAACGCTGGAAATCCGAAATTCAGTACAC 4305
QY 500 rSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrTh 520
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4306 TTCCAACTACAAACAGTCTGTTAATCGTGGACTTACCGTGGATACTAATGGCGTGATTC 4365
QY 520 rGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4366 AGAGCCTCGCCCATTTGGCACCAGATACCTGACTCGTAATCTG 4408
```

Search completed: February 14, 2004, 17:23:51
Job time : 475.286 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 14, 2004, 16:11:58 ; Search time 78.8571 Seconds
(without alignments)
2988.931 Million cell updates/sec

Title: US-09-807-802A-17

Perfect score: 2906

Sequence: 1 MASGGGAPWADNNEGADGVG.....NNGLYTEPRPTGTRYLTRPL 534

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09807802/runat.11022004.175609.15955/app.query.fasta_1.2389
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOKCL=0
-LOOPEXT=0 -UNITS=BITS -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09807802 @CIGN 1 131 @runat.11022004.175609.15955 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------------|
| 1 | 2518 | 86.6 | 8179 | 4 | US-09-438-268-5 |
| 2 | 2481.5 | 85.4 | 4072 | 4 | US-09-770-315-4 |
| 3 | 2481.5 | 85.4 | 7557 | 4 | US-09-770-315-3 |
| 4 | 2481.5 | 85.4 | 8698 | 4 | US-09-770-315-2 |
| 5 | 2439.5 | 83.9 | 4680 | 1 | US-08-254-358-1 |
| 6 | 2439.5 | 83.9 | 4680 | 1 | US-08-475-391-1 |
| 7 | 2439.5 | 83.9 | 4680 | 2 | US-08-709-609-1 |
| 8 | 2439.5 | 83.9 | 4680 | 5 | PCT-US95-07178-1 |
| 9 | 2258.5 | 77.7 | 8151 | 4 | US-09-438-268-2 |
| 10 | 1690.5 | 58.2 | 1617 | 4 | US-09-532-594B-19 |
| 11 | 1690.5 | 58.2 | 1800 | 4 | US-09-532-594B-17 |
| 12 | 1690.5 | 58.2 | 2208 | 4 | US-09-532-594B-5 |

| | | | | | | |
|----|--------|------|------|---|--------------------|--------------------|
| 13 | 1690.5 | 58.2 | 4767 | 4 | US-09-532-594B-1 | Sequence 1, Appli |
| 14 | 1690.5 | 58.2 | 7214 | 4 | US-09-438-268-1 | Sequence 1, Appli |
| 15 | 491.5 | 16.9 | 2271 | 4 | US-09-438-268-3 | Sequence 3, Appli |
| 16 | 238 | 8.2 | 1740 | 1 | US-07-969-213-1 | Patent No. 5223424 |
| 17 | 212 | 7.3 | 1820 | 6 | 5223424-12 | Patent No. 5223424 |
| 18 | 209.5 | 7.2 | 2062 | 6 | 5223424-3 | Sequence 1, Appli |
| 19 | 194 | 6.7 | 2254 | 1 | US-08-552-369-1 | Sequence 1, Appli |
| 20 | 189 | 6.5 | 5049 | 1 | US-08-336-345-1 | Sequence 2, Appli |
| 21 | 189 | 6.5 | 5049 | 2 | US-08-336-345-2 | Sequence 1, Appli |
| 22 | 189 | 6.5 | 5049 | 1 | US-08-647-655-1 | Sequence 2, Appli |
| 23 | 189 | 6.5 | 5049 | 2 | US-08-647-655-2 | Sequence 1, Appli |
| 24 | 183 | 6.3 | 1752 | 3 | US-09-022-949-1 | Sequence 1, Appli |
| 25 | 178 | 6.1 | 1755 | 2 | US-08-317-785-1 | Sequence 201, App |
| 26 | 119 | 4.1 | 3288 | 4 | US-09-107-532A-201 | Sequence 7, Appli |
| 27 | 112.5 | 3.9 | 2394 | 3 | US-09-319-989-7 | Sequence 25, Appl |
| 28 | 111 | 3.8 | 2000 | 2 | US-08-469-202-25 | Sequence 32, Appl |
| 29 | 111 | 3.8 | 2000 | 4 | US-08-484-434C-32 | Sequence 32, Appl |
| 30 | 111 | 3.8 | 2000 | 4 | US-09-384-361-32 | Sequence 3, Appli |
| 31 | 109 | 3.8 | 3561 | 1 | US-08-485-568A-3 | Sequence 5, Appli |
| 32 | 109 | 3.8 | 3561 | 2 | US-08-357-698-5 | Sequence 3, Appli |
| 33 | 109 | 3.8 | 3561 | 2 | US-08-590-554A-3 | Sequence 3, Appli |
| 34 | 109 | 3.8 | 3561 | 2 | US-09-184-223-3 | Sequence 5, Appli |
| 35 | 109 | 3.8 | 3561 | 5 | PCT-US93-12682-5 | Sequence 5, Appli |
| 36 | 108 | 3.7 | 1988 | 1 | US-08-469-202-26 | Sequence 26, Appl |
| 37 | 108 | 3.7 | 1988 | 2 | US-08-484-434C-33 | Sequence 33, Appl |
| 38 | 108 | 3.7 | 1988 | 4 | US-09-384-361-33 | Sequence 33, Appl |
| 39 | 106 | 3.6 | 3155 | 2 | US-08-231-193A-13 | Sequence 13, Appl |
| 40 | 106 | 3.6 | 3155 | 2 | US-08-486-273A-13 | Sequence 13, Appl |
| 41 | 106 | 3.6 | 3155 | 3 | US-08-480-474-13 | Sequence 13, Appl |
| 42 | 106 | 3.6 | 3155 | 3 | US-08-940-086A-13 | Sequence 13, Appl |
| 43 | 106 | 3.6 | 3155 | 4 | US-08-940-035A-13 | Sequence 13, Appl |
| 44 | 106 | 3.6 | 3155 | 4 | US-08-935-105A-13 | Sequence 13, Appl |
| 45 | 106 | 3.6 | 3155 | 4 | US-09-648-797-13 | Sequence 13, Appl |

ALIGNMENTS

RESULT 1

US-09-438-268-5

; Sequence 5, Application US/09438268

; Patent No. 6491907

; GENERAL INFORMATION:

; APPLICANT: Rabinowitz, Joseph E.

; APPLICANT: Samulski, Richard J

; APPLICANT: Xiao, Weidong

; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING

; FILE REFERENCE: 5470-186

; CURRENT APPLICATION NUMBER: US/09/438,268

; CURRENT FILING DATE: 1999-11-10

; EARLIER APPLICATION NUMBER: 60/107,840

; EARLIER FILING DATE: 1998-11-10

; EARLIER APPLICATION NUMBER: 60/123,651

; EARLIER FILING DATE: 1999-03-10

; NUMBER OF SEQ ID NOS: 59

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 5

; LENGTH: 8179

; TYPE: DNA

; ORGANISM: Virus

US-09-438-268-5

Alignment Scores:
Pred. No.: 4,918-275 Length: 8179
Score: 2518.00 Matches: 458
Percent Similarity: 90.65% Conservative: 27
Best Local Similarity: 85.61% Mismatches: 48
Query Match: 86.65% Indels: 2
DB: 4 Gaps: 2

US-09-807-802A-17 (1-534) x US-09-438-268-5 (1-8179)

QY 1 MetalAserGlyGlyGlyAlaProMetAlaAspAsnGluGlyAlaAspGlyValGly 20

Db 2739 ATGGCTACAGGAGTGGCGCCACCAATGCGAGACATACAGAGGCGCGAGCGAGTGGGT 2798
Qy 21 AenAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 40
Db 2799 AATTCCTCGGAAATGGCAATGGCAATTCACATGGATTCACATGGGCGACAGAGTCATCACACC 2858
Qy 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrIleGlnIleSerSer 60
Db 2859 AGCACCCGGAACCTGGCGCTGCGCCACCTACAAACACCACTCTACAAACAAATTTCCAGC 2918
Qy 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
Db 2919 CAATCA--GGAGCTCGAAGCAGCAATCACTACTTTGGCTACAGCACCCCTTTGGGGGTAT 2975
Qy 81 PheAppPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
Db 2976 TTTGACTTCAACAGATTCACATGCCACTTTTCCACACGTCGACTGGCAAGACTCATCAAC 3035
Qy 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLeuPheAsnIleGlnValLys 120
Db 3036 AACAACTGGGGATTCGAGCCCAAGAGACTCAACTTCAAGCTCTTTAAACATTCAGTCAAA 3095
Qy 121 GluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeuThrSerThrValGln 140
Db 3096 GAGGTACCCAGATGACGGTACGACGAGCATTTGCCAATAACCTTTACACGACGGTTACG 3155
Qy 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
Db 3156 GTGTTTACTGACTCGGAGTACCAAGCTCCCGTACGTCTCGGGTCGGCGCACCAAGCTGT 3215
Qy 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
Db 3216 CTCCCCCGCTTCCAGCGAGAGTCTTCATGGTCCCTCAGTATGGATACCTCACCCCTGAAAC 3275
Qy 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
Db 3276 AACGGAAGTCAAGCGTGGGAGCGCTCATCTTTTACTGCTGGAGTACTTCCCTTCGCGAG 3335
Qy 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
Db 3336 ATGCTAAGGACTCGGAATTAATTTCCAAATTCAGCTATACCTTCGAGGATGACTTTTCAC 3395
Qy 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
Db 3396 AGCAGCTACGCTCACAGCCAGAGTTGGATCGCTTGATGATGATCTCTTATTGATCAGTAT 3455
Qy 241 LeuTyrTyrLeuAsnArgThrGln---AsnGlnSerGlySerAlaGlnAsnLysAspLeu 259
Db 3456 CTGTACTTACTGAACAGACGAGCAAGCAACCTCTGGAACCAACCAACCAATCACGGCTG 3515
Qy 260 LeuPheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGly 279
Db 3516 CTTTATAGCCAGCGCTGGCGCTTCAGTCTATGCTTTTGGCGCCAGAAATTTGGCTACCTGG 3575
Qy 280 ProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPhe 299
Db 3576 CCTGTCTACCGGCACAGAGACTTTCAAAGACTGCTTAACGACCAACCAACCACTAATTT 3635
Qy 300 ThrTrpThrGlyAlaSerTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGly 319
Db 3636 CCTTGGACAGCGCCAGCAAAATATCATCTCAATGGCGCGACTCGCTGGTGAATCCAGGA 3695
Qy 320 ThrAlaMetAlaSerHisLysAspAspGluAspLysPhePheMetSerGlyValMet 339
Db 3696 CCAGCTATGGCCAGTCAACAGGACGATGAAGAAATTTTTCCTATGACGCGCAATCTA 3755
Qy 340 IlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThr 359
Db 3756 ATATTGGCAAGAGGACCAACGGCAAGTAAACGAGAAATAGATATGTAATGATACG 3815
Qy 360 AspGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAla 379

Db 3816 GATGAAGAAGAGATTCTGATCCACCACCAATCTCTGTGGCAACAGAGCAGTAGTATGGAACTGTGGCA 3875
Qy 380 ValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyVala 399
Db 3876 AATAAAGTTGGAGAGCTCAATATACAGCTCCCAACAGCTGGAATGTCAATCATCAGGGGGGCC 3935
Qy 400 LeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLys 419
Db 3936 TTACTCTGCATGGTGTGGCAAGATCGTGACGTGTACCTTCAAGGACCTATCTGGGCAAG 3995
Qy 420 IleProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLys 439
Db 3996 ATTCTCACACGAGTGGACACTTTTCATCTCTCTCTGTGGAGGCTTTGGACTGAAA 4055
Qy 440 AsnProProProGlnIleLeuIleLysAsnThrProValProAlaAsnProAlaGlu 459
Db 4056 CATCGGCTCTCAATATCATGATCAAAATTAATCTCGGTACTCTCGAATCTCTTCACACC 4115
Qy 460 PheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerVal 479
Db 4116 TTCAAGTCGGCAAGTTTGTCTTCTTCATCACAGTACTCCACGGGACAGGTCAAGCGT 4175
Qy 480 GluIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyr 499
Db 4176 GAGATCGAGTGGGAGCTGCAGAGGAAACAGCAACCGCTGGAATCCCGAAATTCAGTAC 4235
Qy 500 ThrSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyr 519
Db 4236 ACTTCCAACTACAACAGTCTGTTAATCGTGGACTTACCGTGGATACTAATGGCGTGTAT 4295
Qy 520 ThrGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
Db 4296 TCAGAGCTCGCCCATTTGGCACCAGATACCTGACTCGTAATCTG 4340

RESULT 2

US-09-770-315-4
; Sequence 4, Application US/09770315
; Patent No. 6429001
; GENERAL INFORMATION:
; APPLICANT: Chiron Corporation
; TITLE OF INVENTION: Recombinant AAV Packaging Systems
; FILE REFERENCE: 20263-501
; CURRENT APPLICATION NUMBER: US/09/770,315
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,536
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: recombinant DNA
US-09-770-315-4

Alignment Scores:
Pred. No.: 2,16e-271 Length: 4072
Score: 2481.50 Matches: 445
Percent Similarity: 90.28% Conservative: 37
Best Local Similarity: 83.33% Mismatches: 51
Query Match: 85.39% Indels: 1
DB: 4 Gaps: 1

US-09-807-802A-17 (1-534) x US-09-770-315-4 (1-4072)

Qy 1 MetaLaserGlyGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGly 20
Db 2090 ATGGCTACAGGAGTGGCGCCACCAATGGCAGCAATTAACAGAGGCGCCACGAGTGGGT 2149
Qy 21 AnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 40
Db 2150 AATTCCTCGGAAATTTGGCAATTTGGCAATTTCCATGATGGGCGACAGAGTCATCACACC 2209

QY 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSer 60
DB 2210 AGCACCCGAACTGGGCCCTGCCACCTACAAACACCACTCTACAAACAAATTTCCAGC 2269
QY 61 AlaSerThrGlyAlaSerAsnAsnHisTyrPheGlyTyrSerThrProTyrGlyTyr 80
DB 2270 CAATCA---GGAGCCTCGAACACAAATCACTACTTTGGCTACAGCACCCCTGGGGGTAT 2326
QY 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
DB 2327 TTTGACTTCAACAGATTCCACTGCCACCTTTTCCACCGTGAGCTGGCAAGACTCATCAAC 2386
QY 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 120
DB 2387 AACAACTGGGGATTCCGACCCCAAGAGACTCAACTTCAAGCTCTTTAACTCAAGTCAAA 2446
QY 121 GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 140
DB 2447 GAGGTACGCGAGATGACGGTACGACGAGATTGCGCAATAACCTTACCAGCACGGTTCCAG 2506
QY 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
DB 2507 GTGTTTACTGACTCCGAGTACCGACTCCCGTACGTCTCGGCTCGCGCGCATCAAGGATGC 2566
QY 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
DB 2567 CTCCTCCGCGTTCCTCCAGCAGAGCTTCTATGGTGCCACAGTATGGATACCTCACCCCTGAAC 2626
QY 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
DB 2627 AACGGGAGTACGGCAGTAGGAGCGCTCTTCACTTTTACTGCTCGCGAGTACTTCTCTCTCAG 2686
QY 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
DB 2687 ATGCTGCGTACCGGAAACAACTTACCTTACGTACACTTTTGGAGAGTTCCTTTCCAC 2746
QY 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
DB 2747 AGCAGCTACGCTCACAGCCAGAGTCTGGACCGTCTCATGAATCTCTCTCATCGACCGATAC 2806
QY 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
DB 2807 CTGTATTACTTGGACGACAAACAACTCCAAAGTGGAAACCAACCCAGCTCAAGGCTTCAG 2866
QY 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
DB 2867 TTTTCTACGGCCGAGCGAGTGACATTCCGGACCACTCTAGGAAGTCTCTCTGGACCC 2926
QY 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
DB 2927 TGTACCGCCAGCAGCGAGTATCAAAAGACATCTGGCGGATACAAACAAACAGTGAATACTCG 2986
QY 301 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr 320
DB 2987 TGGACTGAGCTTCAACAGTACCACTCAATGGCAGAGACTCTCTGGTGAATCCCGGCCCG 3046
QY 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
DB 3047 GCCATGGCAGCCACAGGACGATGAGAGAAAGTTTTTCTTCAGCGGGGGTCTCATC 3106
QY 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
DB 3107 TTTGGGAAGCAAGGCTCAGAGAAAACAAATGTGGACATTGAAAAGGTCTATGATTACAGAC 3166
QY 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
DB 3167 GAAGAGAAATTCAGCAACCACTCCCGTGGCTACGGAGCGATGTGTTCTGTATCTACC 3226
QY 381 AsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
DB 3227 AACCTCAGAGGCAACAGACACAGCAGTACCGCAGATGTCAACACACAAAGGCGTCTTT 3286

QY 401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 420
DB 3287 CCAGGCATGGTCTGGCAGGACAGAGATGTGTACTCTCAGGGGCCCATCTGGGCAAGATT 3346
QY 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
DB 3347 CCACACACGACGAGCAATTTTCCACCTCTCCCTCATGGGTGGATTGGGACTTAAACAC 3406
QY 441 ProProGlnIleLeuIleLysAsnThrProValProAlaAsnProAlaGluPhe 460
DB 3407 CCTCTCCACAGATTCTCATCAAGAACACCCCGGTACTTGGCAATCTCTGCACCACTTC 3466
QY 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
DB 3467 AGTCGGCAAGTTTGTCTCTTTCATCACAGTACTCCACGGACAGGTCCAGCGTGGAG 3526
QY 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrThr 500
DB 3527 ATCGAGTGGGAGCTGCGAGAGGAAACAGCAACCGCTGGAAATCCCGAAATTCAGTACT 3586
QY 501 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 520
DB 3587 TCCAACTACAAAGTCTGTAATGTGGACTTTTACTGTGGACACTAATGGCGGTATTCA 3646
QY 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
DB 3647 GAGCTCGCCCAATTGGCACCAGATACCTGACTCGTAATCTG 3688
RESULT 3
US-09-770-315-3
; Sequence 3, Application US/09770315
; Patent No. 6429001
; GENERAL INFORMATION:
; APPLICANT: Chiron Corporation
; TITLE OF INVENTION: Recombinant AAV Packaging Systems
; FILE REFERENCE: 20263-501
; CURRENT APPLICATION NUMBER: US/09/770,315
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,536
; PRIOR FILING DATE: 2000-01-26
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 7557
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: recombinant DNA
US-09-770-315-3
Alignment Scores:
Pred. No.: 6,09e-271 Length: 7557
Score: 2481.50 Matches: 445
Percent Similarity: 90.26% Conservative: 37
Best Local Similarity: 83.33% Mismatches: 51
Query Match: 85.39% Indels: 1
DB: 4 Gaps: 1
US-09-807-802A-17 (1-534) x US-09-770-315-3 (1-7557)
QY 1 MetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGly 20
DB 2837 ATGGCTACAGCGAGTGGCGCGCAATATGGCAGCAATATACGAGGCGCGCGAGTGGGT 2896
QY 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 40
DB 2897 AATTCCTCGGAAATTTGGCATTTCCATGGATGGCGGACAGAGTCTATCCACC 2956
QY 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSer 60
DB 2957 AGCACCCGAACTGGGCGCTGCCACCTTACAAACCACTCTTACAAACAAATTTCCAGC 3016
QY 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTyrGlyTyr 80

Db 3017 CAATCA---GGAGCCTCGAAGCAGCAATCACTTTGGCTACAGACCCCTTCGGGGTAT 3073
Qy 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
Db 3074 TTTGACTTCAACAGATTCCACTGCGCACTTTTACCACGTCGACGTGGCAAGACTCATCAAC 3133
Qy 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPhelysLeuPheAsnIleGlnValLys 120
Db 3134 AACAACTGGGGATTCCGACCCAGAGACTCAACTTCAAGCTCTTTAACTTCAAGTCAAA 3193
Qy 121 GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 140
Db 3194 GAGGTACGCGAGATGACGCTACGACGAGCTGCTTCCATTAACCTTACCACGAGTCAG 3253
Qy 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
Db 3254 GTGTTTACTGACTCGGAGTACAGCTCCCGTACGCTCGGCTCGGCGCATCAAGATGC 3313
Qy 161 LeuProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
Db 3314 CTCCCGCGCTTCCAGCAGAGCTCTTATGCTGTCACAGTATGATGATCACTCACCCCTGAAC 3373
Qy 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
Db 3374 AACGGAGTCAGGAGTAGAGCTCTTCACTTACTGCTGGAGTACTTCTCTTCAG 3433
Qy 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
Db 3434 ATGCTGCTGACCGGAAACAACTTTTACCTTCAGCTACACTTTTGGAGGAGCTTCTCTCCAC 3493
Qy 221 SerSerTyrAlaHisSerGlnSerLeuAsnArgLeuMetAsnProLeuIleAspGlnTyr 240
Db 3494 AGCAGCTACGCTACAGCCAGAGTCTGGACCGTCTCATGAATCTCTCATCGACCACTAC 3553
Qy 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeu 260
Db 3554 CTGTATTATTGACGACAGCAACAACTCCAACTGGAACCAACCAACAGCTCAAGCTTCAG 3613
Qy 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
Db 3614 TTTTCTCGCGCGGAGGAGTGAGCAATTCGGGACCACTGCTAGGAACCTGCTCTCGGACCC 3673
Qy 281 CysTyrArgGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
Db 3674 TGTTACCGCCAGCAGCAGGATCAAGACATCTCGGATACAAACAAACAGTGAATCTCG 3733
Qy 301 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleLeuAsnProGlyThr 320
Db 3734 TGGACTGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTGTGTGAATTCGGGCCG 3793
Qy 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
Db 3794 GCCATGGCAGCCCAAGGAGCATGAAGAAAAGTTTTTCTCAGAGCGGGTTCTCATC 3853
Qy 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
Db 3854 TTTGGGAAGCAAGCTCAGAAAACAAATGTGCAATTTGAAAGGTCAATGATTAACAGAC 3913
Qy 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
Db 3914 GAACAGGAAATCAGGACAAACCAATCCGTGGCTACGGAGCAGTATGTTCTGTATCTACC 3973
Qy 381 AsnPheGlnSerSerThrAspProAlaThrLysAspValHisAlaMetGlyAlaLeu 400
Db 3974 AACCTCCAGAGAGGCAACAGACAGTACCCAGATGTCAACACACAAAGCGCTTCTT 4033
Qy 401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 420
Db 4034 CCAGGCATGGTCTGGCAGGACAGATGTGTACCTTCAGGGGCCCATCTCTGGGCAAGATT 4093
Qy 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440

Db 4094 CCACACGCGAGCGAGATTTTCCACCCCTCTCCCTCATGGGTGGATTTCGGACTTAAACAC 4153
Qy 441 ProProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460
Db 4154 CCTCTCCACAGATTCTCATCAAGAACACCCCGGTACTTCGGAATCTCTTCGACCACTTC 4213
Qy 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
Db 4214 AGTGGCGAAGTTTGTCTTCTTCATCACAGTACTCACCGGACAGTTCAGCGTGGAG 4273
Qy 481 IleGluTrpGluLeuGlnLysGluAsnSerLysArgTyrAsnProGluValGlnTyrThr 500
Db 4274 ATCGAGTGGGAGCTGCAGAAAGAAACAGCAACCGCTGGAATCCCGAAATTCAGTACACT 4333
Qy 501 SerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrThr 520
Db 4334 TCCAACTACAACAGTCTGTTAATGTGGACTTTTACTGTGGACACTAATGGCGTGTATTCA 4393
Qy 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
Db 4394 GAGCCTCGCCCATTTGGCACCAGATACCTGACTCGTAATCTG 4435

RESULT 4

US-09-770-315-2
; Sequence 2, Application US/09770315
; Patent No. 6429001
; GENERAL INFORMATION:
; APPLICANT: Chiron Corporation
; TITLE OF INVENTION: Recombinant AAV Packaging Systems
; FILE REFERENCE: 20263-501
; CURRENT APPLICATION NUMBER: US/09/770,315
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,536
; PRIOR FILING DATE: 2000-01-26
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 8698
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: recombinant DNA
US-09-770-315-2

Alignment Scores:
Pred. No.: 7,71e-271 Length: 8698
Score: 2481.50 Matches: 445
Percent Similarity: 90.26% Conservative: 37
Best Local Similarity: 83.33% Mismatches: 51
Query Match: 85.39% Indels: 1
DB: 4 Gaps: 1

US-09-807-802A-17 (1-534) x US-09-770-315-2 (1-8698)

Qy 1 MetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGly 20
Db 2809 ATGCTTACAGCGCAGTGGCGCACCAATGGCAGACAATAACGAGGGCGCGCAGGAGTGGT 2868
Qy 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 40
Db 2869 AATTCTCGGAAATTTGGCAATTCGGAATTCGCAATGATGGCGCAGAGTCAATCACACC 2928
Qy 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSer 60
Db 2929 AGCACCAGAACCTGGGCGCTTCCACCTACACACACCACTCTACAAACAAATTTCCAGC 2988
Qy 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
Db 2989 CAATCA---GGAGCCTCGAAGCAGCAATCACTACTTTTGGCTACAGCACCCCTTGGGGTAT 3045
Qy 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
Db 3046 TTTGACTTCAACAGATTCCCACTGCCACTTTTCCACCGTGCAGTGGCAAGAGACTCATCAAC 3105

```
Qy 101 AenAenTrpGlyPheArgProLysArgLeuAenPheLysLeuPheAenIleGlnVallys 120
Db 3106 AACAACTGGGATTCGGACCCAGAGACTCACTTCAAGCTCTTTAACAATTCAGTCAA 3165
Qy 121 GluValThrThrAenAspGlyValThrThrIleAenAenLeuThrSerThrValGln 140
Db 3166 GAGGTACAGCAGATGACGTACGACGACGATTCCTCAATATACCTTTACGACGCGTTACG 3225
Qy 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
Db 3226 GTGTTTACTGACTCGGATACAGCTCCGCTACGCTCCGCTCGGCGCATCAAGGATGC 3285
Qy 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAen 180
Db 3286 CTCGCGCGTTCGCCAGCAGCTCTTCATGGTGCACAGTATGGATACCTCACCTCGAAC 3345
Qy 181 AenGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
Db 3346 AACGGGAGTCAGGAGTAGGACGCTCTTCATTTACTGCTGGAGTACTTTCTCTTCAC 3405
Qy 201 MetLeuArgThrGlyAenAenPheThrPheSerTyrThrPheGluGluValProPheHis 220
Db 3406 ATGCTGCGTACCGGAAACAACCTTTACCTTCAGCTACACTTTTGGAGGAGCTTCTTCAC 3465
Qy 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAenProLeuIleAspGlnTyr 240
Db 3466 AGCAGCTACGCTCACAGCAGAGCTGGACGCTCTCATGAATCTCTCATCGACGATAC 3525
Qy 241 LeuTyrTyrLeuAenArgThrGlnAenGlnSerGlySerAlaGlnAenLysAspLeuLeu 260
Db 3526 CTGTATTACTTTGAGCAGAACAACTCCAAATGGGAACCCAGCCAGCTCAAGGCTTCAG 3585
Qy 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAenTrpLeuProGlyPro 280
Db 3586 TTTTCTCAGCGGAGGAGTGACATTCGGACCACTCTAGAACTGGCTTCTCTGGACCC 3645
Qy 281 CysTyrArgGlnArgValSerLysThrLysThrAspAenAenAenSerAenPheThr 300
Db 3646 TGTTACCGCCAGCAGGAGTATCAAGACATCTCGGGATACCAACAACAGTGAATACCTCG 3705
Qy 301 TrpThrGlyAlaSerLysTyrAenLeuAenGlyArgGluSerIleAenProGlyThr 320
Db 3706 TGGACTGGAGCTACCAAGTACCACTCAATGGCAGAGACTCTCTGTGTAATCCGGGCGCG 3765
Qy 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
Db 3766 GCATGGCAGCCACAGGACCATGAAGAAAGTTTTTCTTCAGAGCGGGTTCTCATC 3825
Qy 341 PheGlyLysGluSerAlaGlyAlaSerAenThrAlaLeuAenAenValMetIleThrAsp 360
Db 3826 TTTGGGAAGCAAGGCTCAGAGAAACAATGTGGACATTTGAAAGGTCTATGATTACAGAC 3885
Qy 361 GluGluGluIleLysAlaThrAenProValAlaThrGluArgPheGlyThrValAlaVal 380
Db 3886 GAAGAGGAATCAGACAAACCAATCCCGTGGCTACGGGAGCATGTGGTTCTGTATCTACC 3945
Qy 381 AenPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
Db 3946 AACCTCCAGAGGACACAGACAGCTACCGAGATGTCAACACACAGGCGCTTCTT 4005
Qy 401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 420
Db 4006 CCAGGCATGTTCTGGCAGGACAGAGATGTGTACCTTCAGGGGCCCATCTGGGCAAGATT 4065
Qy 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyLysPheGlyLeuLysAen 440
Db 4066 CCACACGGACGGACATTTTCACCCCTCTCCCTCATGGGTGGATTCGGACTTAAACAC 4125
Qy 441 ProProProGlnIleLeuLysAenThrProValProAlaAenProAlaGluPhe 460
Db 4126 CCTCTCCACAGATTCTCATCAAGAACACCCCGGTACCTGCGAATCTCTTCGACCACTTC 4185
```

```
Qy 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGlnValSerValGlu 480
Db 4186 AGTGGGCAAGTTTGCTTCTTCATCACACAGTACTCCACGGGACAGGTCCAGCGTGGAG 4245
Qy 481 IleGluTrpGluLeuGlnLysGluAenSerLysArgTrpAenProGluValGlnTyrThr 500
Db 4246 ATCGAGTGGGAGCTGCAAGAGGAAACAGCAACGCTGGAATCCCGAAATTCAGTACACT 4305
Qy 501 SerAenTyrAlaLysSerAlaAenValAspPheThrValAenAenAenGlyLeuTyrThr 520
Db 4306 TCCAACTACAAAGTCTGTATATGGACTTTTACTGTGGACACTAATGGCGTGTATCA 4365
Qy 521 GluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
Db 4366 GAGCTCGCCCATTTGGCACCAGATACCTGACTCGTAATCTG 4407

RESULT 5
US-08-254-358-1
; Sequence 1, Application US/08254358
; Patent No. 5658785
; GENERAL INFORMATION:
; APPLICANT: Johnson, Philip R.
; TITLE OF INVENTION: Adeno-Associated Virus Materials and
; METHOD OF INVENTION: Methods
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/254,358
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5658785and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31975
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4680 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-254-358-1

Alignment Scores:
Pred. No.: 1.63e-266 Length: 4680
Score: 2439.50 Matches: 441
Percent Similarity: 89.16% Conservative: 36
Best Local Similarity: 82.43% Mismatches: 56
Query Match: 83.95% Indels: 2
DB: 1 Gaps: 1

US-09-807-802A-17 (1-534) x US-08-254-358-1 (1-4680)

Qy 1 MetAlaSerGlyGlyGlyAlaProMetAlaAenAenGluGlyAlaAenGlyValGly 20
Db 2809 ATGGCTACAGGCACTGGCGCACCAATTCGACACAATAACGAGGCGCGGAGTGGGT 2868
Qy 21 AsnAlaSerGlyAenTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 40
```

Db 2869 AATTCCTCCGAAATGGCAATTCGATTCACATGGGCGACAGAGTCATACCAACC 2928
QY 41 SerThrArgThrTrrAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSer 60
Db 2929 AGCACCCGAACTGGGCGCTGCCACCTACACACCAACCACTCTACAAACAATTTCCAGC 2988
QY 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrrpGlyTyr 80
Db 2989 CAATCA--GGAGCCTCGAACGACAACTACTATTGGCTACAGCACCCCTTTGGGGGTAT 3045
QY 81 PheAspPheAsnArgPheHisCysHisPheSerProAspAspTrrGlnArgLeuIleAsn 100
Db 3046 TTGAGCTTCAACAGATTCCACTGCCACTTTTCCACACGGTACTGGCAAGACTCATCAAC 3105
QY 101 AsnAsnTrrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 120
Db 3106 AACAACTGGGGATTCCGACCCCAAGAGACTCAACTTCAAGCTCTTTAAACATTCAAGTCAAA 3165
QY 121 GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 140
Db 3166 GAGGTCAACAGAAATGACGGTACGACGAGATTGGCCAAATAACCTTTACCAAGACGGTTACG 3225
QY 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
Db 3226 GTGTTTACTGACTCGGAGTACAGCTCCCGTACGCTCCGCTCGGCGCATCAAGATGC 3285
QY 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTrrLeuThrLeuAsn 180
Db 3286 CTCCCGCGCTTCCAGCAGACGCTTTCATGGTGCCACAGTATGATACCTCAACCTGAAC 3345
QY 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
Db 3346 AACGGAGTCAGCAGTAGGACGCTCTTCAATTTACTGCTGGAGTACTTCTCTCTCAG 3405
QY 201 MetLeuArgThrGlyValAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
Db 3406 ATGCTCGGTACCCGGAACAACTTTTACTCTCAGTACACTTTTGGAGCGCTTCTTCCAC 3465
QY 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
Db 3466 AGCAGCTACGCTCACAGCCAGAGTCTGGACCGTCTCATGAATCTCTCATCGACCAAGTAC 3525
QY 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
Db 3526 CTGTATTACTTGGACGAGCAAAACACTCCCAAGTGGAAACCAACCAAGCGTCAAGGCTTCAG 3585
QY 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrrpLeuProGlyPro 280
Db 3586 TTTTCTCAGCCCGGAGGTGATCTCGGACCACTTAGGNACTGGCTTCTTGGACCC 3645
QY 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
Db 3646 TGTACCGCCAGCAGCGAGTATCAAGACATCTCGCGATAACAAACAGTGAATCTCG 3705
QY 301 TrrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgLysSerIleLeuAsnProGlyThr 320
Db 3706 TGGACTGGAGTACCAGTACCACTCAATGGCAGAGACTCTCTGTGTGAATCCGGGGCCC 3765
QY 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
Db 3766 GCATGGCAGCCACCAAGGACGATGAAGAAAGTTTTCCTCAGAGCGGGTCTCATC 3825
QY 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
Db 3826 TTTGGGAAGCAGGCTCAGAGAAACAAATGTGAACATGAAAGGTGATGATTACAGAC 3885
QY 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
Db 3886 GAAGAGGAATCGGAACAACTCCCGTGGCTACGAGCAGTATGTTCTGTATCTACC 3945
QY 381 AsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400

Db 3946 AACCTCCAGAGAGCAACAGACAGCTACCGCAGATGTCAACACACACAGGCGTTCTT 4005
QY 401 ProGlyMetValTrrpGlnAspArgAspValTyrLeuGlnGlyProIleTrrpAlaLysIle 420
Db 4006 CCAGCATGGTCTGGCAGGACAGAGATGTGTACCTTCAGGGGCGCCATCTGGGCAAGATT 4065
QY 421 ProIleThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
Db 4066 CCACACAGGACGAGCATTTTCCACCCCTCTCCCTCATGGGTGGATTCCGACTTAAACAC 4125
QY 441 ProProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460
Db 4126 CCTCTCCACAGATTCTCATCAAGACACACCCCGGTACCTCGGAATCTCTTCACCACTTC 4185
QY 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGln-ValSerValGln 480
Db 4186 AGTGGCGCAAGTTTGGTCTTCTTCATCACAGTACTCCACGGGACACGGTCAGCGTGA 4245
QY 480 uileGluTrrpGluLeuGlnLysGluAsnSerLysArgTrrpAsnProGluValGlnTyrTh 500
Db 4246 GATCGAGTGGGAGCTGCAGAAAGGAAACAGCAACGCTGGAAATCCGGAATTCAGTACAC 4305
QY 500 rSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrTh 520
Db 4306 TTCCAACTACAAAGTCTGTAAATCGTGGACTTACCGTGGATACTAATGGCGTGTATTC 4365
QY 520 rGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
Db 4366 AGAGCTCGCCCATTTGGCACCAGATACCTGACTCGTAATCTG 4408

RESULT 6

US-08-475-391-1
; Sequence 1, Application US/08475391
; Patent No. 5786211
; GENERAL INFORMATION:
; APPLICANT: Johnson, Philip R.
; TITLE OF INVENTION: Adeno-Associated Virus Materials and
; TITLE OF INVENTION: Methods
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,391
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/254,358
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5786211and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31975
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4680 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

US-08-475-391-1

Alignment Scores:

Pred. No.: 1,63e-266 Length: 4680
Score: 2439.50 Matches: 441
Percent Similarity: 89.16% Conservative: 36
Best Local Similarity: 82.43% Mismatches: 56
Query Match: 83.95% Indels: 2
DB: 1 Gaps: 1

US-09-807-802A-17 (1-534) x US-08-475-391-1 (1-4680)

QY 1 MetAlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGly 20
DB 2809 ATGGCTACAGGAGTGGCGCACCAATGGCAGACAATAACAGAGGCGCGAGTGGGT 2868
QY 21 AsnAlaSerGlyAsnTrpHisCysValAspSerTrpLeuGlyAspArgValIleThr 40
DB 2869 AATTCTCCGGAATTTGGCAATTTGCCAATTCACATGGATGGCGCAGAGTCAATCACCAC 2928
QY 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSer 60
DB 2929 AGCACCCGAACCTGGCGCCCTGCCACCTACACCAACACCTCTCAACAAATTTCCAGC 2988
QY 61 AlaSerThrGlyAlaSerAsnAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
DB 2989 CAATCA--GGAGCCTCGAACACCAATCACTACTTTGGCTACAGCACCCTTTGGGGGTAT 3045
QY 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
DB 3046 TTTGACTTCAACAGATTCCACTGCCACTTTTCCACCGTGCAGTGCAGAAAGACTCATCAAC 3105
QY 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLeuPheAsnIleGlnValLys 120
DB 3106 AACAACTGGGGATTCGACCCCAAGAGACTCAACTTCAAGCTCTTTAACTCAAGTCAAA 3165
QY 121 GluValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGln 140
DB 3166 GAGGTTCACGACAGATGACGGTACGACGAGCTTTCACCAATTAACCTTTACAGCAGCGTTT 3225
QY 141 ValPheSerAspSerGlnTrpGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
DB 3226 GGTGTTTACTGATCGAGTACAGCTCCGTCAGTCCGCTCGCTCGCGCATCAGAGATGC 3285
QY 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180
DB 3286 CTCCTCCCGCTTCCAGCAGAGCTTCTATGGTGCCACAGATGATGATCACTCACCTCGAAC 3345
QY 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGlyTyrPheProSerGln 200
DB 3346 AACGGGAGTACAGGAGTACAGGAGCTTCTTCAATTTACTGCTGGAGTACTTCTCTTCAG 3405
QY 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
DB 3406 ATGCTGCGTACCGGAACCACTTTTACCTTCAGTACACTTTTGGAGAGCTTCTTTCAC 3465
QY 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
DB 3466 AGCAGTACGCTCACAGCCAGAGTCTGGACCGCTCTCATGAATCTCTCATCGACCAGTAC 3525
QY 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnIleAspLeu 260
DB 3526 CTGTATTACTTGAAGACCAACAACTCCAAGTGAACCAACCCACGACGATCAAGGCTTCAG 3585
QY 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
DB 3586 TTTTCTCAGCGCGGAGGTGACATTCGGGACGAGTCTAGGAATCGCTTCTTCCGACCC 3645
QY 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnSerAsnPheThr 300
DB 3646 TGTTACCGCCAGCAGGAGTATCAAGACATCTCGGGATACCAACCAACAGTGAATACTCG 3705
QY 301 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr 320

DB 3706 TGGACTGGAGCTCAAGATACCACTCAATGGCAGAGACTCTCTGTGTAATCCGGGGCCC 3765
QY 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
DB 3766 GCCATGGCAAGCCCAAGGACGATGAAGAAANGTTTTTCTTCAGAGCGGGTCTTCATC 3825
QY 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
DB 3826 TTTGGGAAGCAAGGCTCAGAGAAAACAAATGTGAACATTTGAAAGGTTCATGATTACAGAC 3885
QY 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
DB 3886 GAAGAGGAAATTCGGAACCAACCAATCCCGTGGCTACGGAGCATATGGTCTGTATCTACC 3945
QY 381 AsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
DB 3946 AACCTCCAGAGAGCAACAGACAGACTACCCAGATGTCAACACACAGGCGTCTT 4005
QY 401 ProGlyMetValTrpGlnAspArgValTyrLeuGlnGlyProIleTrpAlaLysIle 420
DB 4006 CCAGGCTGGTCTGGCAGGACAGAGATGTGTACCTTCAGGGGCCCATCTGGGCAAGATT 4065
QY 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
DB 4066 CCACACCGGACGACATTTTCACTCCCTCTCCCTCATGGGTGGATTGGGACTTTAAACAC 4125
QY 441 ProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460
DB 4126 CCTCTCCACAGATCTCATCAAGAACACCCCGGTACTTGGGATCTTCACACCATCTTC 4185
QY 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGln-ValSerValGI 480
DB 4186 AGTGGCGCAAGTTTGTCTTCTTCATCACAGTACTCCACGGGACACGGTCAGCGTGA 4245
QY 480 uIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrTh 500
DB 4246 GATCGAGTGGGAGCTGCAGAGGAAACAGCAACCGTGGAAATCCCGAAATTCAGTACAC 4305
QY 500 rSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrTh 520
DB 4306 TTCCACTACACAGTCTGTATTCGTGGACTTACCGTGGATCTAATGCGGTGATTTC 4365
QY 520 rGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
DB 4366 AGAGCTCGCCCATTTGGCACCAGATACCTGACTCGTAATCTG 4408

RESULT 7

US-08-709-609-1
; Sequence 1, Application US/08709609
; Patent No. 5858775
; GENERAL INFORMATION:
; APPLICANT: Johnson, Philip R.
; TITLE OF INVENTION: Adeno-Associated Virus Materials and Methods
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/709,609
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```

; NAME: No. 5858775and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31975
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4680 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-709-609-1

Alignment Scores:
Pred. No.: 1 63e-266 Length: 4680
Score: 2439.50 Matches: 441
Percent Similarity: 89.16% Conservative: 36
Best Local Similarity: 82.43% Mismatches: 56
Query Match: 83.95% Indels: 2
DB: 2 Gaps: 1

US-09-807-802a-17 (1-534) x US-08-709-609-1 (1-4680)

Qy 1 MetAlaSerGlyGlyAlaProMetAlaAspAsnGluGlyAlaAspGlyValGly 20
Db ATGGCTACAGCGAGTGGCGCACCACCAATGGCAGACAATAACGAGGCGCGCAGGAGTGGGT 2868
Qy 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThr 40
Db AATTCCTCCGGAATTTGGCATTCGGATTCCACATGATGGATGGCGACAGATCATCACACC 2928
Qy 41 SerThrArgThrTrpAlaLeuProThrTrpAsnAsnHisLeuTrpLysGlnIleSerSer 60
Db AGCACCCGACCTGGGCGCTGCCACCTACACCAACACCCCTCTACAAACAAATTTCCAGC 2988
Qy 61 AlaSerThrGlyAlaSerAsnAspAsnHisCysPheGlyTrpSerThrProTrpGlyTrp 80
Db CAATCA---GGAGCCTCGAAGCAGCAATCACTACTTTGGCTTACAGCACCCCTTGGGGGTAT 3045
Qy 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
Db TTTGACTTCAACAGATTCACCTGCCACTTTTACACAGTACTGGCAAGACATCATCAAC 3105
Qy 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLys 120
Db AACAACTGGGGAATTCGACCCCAAGAGACTCAACTTCAAGCTCTTTAAACATTTCAAGTCAA 3165
Qy 121 GluValThrThrAsnAspGlyValThrThrIleAlaAsnLeuThrSerThrValGln 140
Db GAGGTACGACGAGATGACGGTACGACGACGATTCGCCAATAACCTTACACGACGGTTTCAG 3225
Qy 141 ValPheSerAspSerGluTrpGlnLeuProTrpValLeuGlySerAlaHisGlnGlyCys 160
Db GTGTTTACTGACTCGGAGTACCACTCCCGTACGTCTCGCTCGGCGCATCAAGATGC 3285
Qy 161 LeuProProPheProAlaAspValPheMetIleProGlnTrpGlyTrpLeuThrLeuAsn 180
Db CTCCCGCGGTTTCCGACGACAGCTCTTCATGTGTGTCACAGATGATGATACCTCACCCCTGAAC 3345
Qy 181 AsnGlySerGlnAlaValGlyArgSerSerPheTrpCysLeuGluTrpPheProSerGln 200
Db AACGGGAGTTCAGGAGTACGAGTCTCTCTATTTTACTGCCTGGAGTACTTTCTCTCTTCAG 3405
Qy 201 MetLeuArgThrGlyAsnAsnPheThrPheSerThrThrPheGluGluValProPheHis 220
Db ATGCTGGGTACCGGAACAACATTTACCTTCAGCTACACTTTTGGAGACGTTCTCTTCCAC 3465
Qy 221 SerSerTrpAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTrp 240
Db AGCAGCTACGCTCACAGCCAGAGTCTGACCGCTCTCATGAATCTCTCTCATCAGCACGATC 3525

```

```

Qy 241 LeuTrpTrpLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeu 260
Db CTGTATTACTTTGAGCAGAAACAACACTCAAGTTGGAACACACACGACGATCAAGGTTCTCAG 3585
Qy 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
Db TTTTCTAGGCGGAGCGAGTGACATTCGGGACCACTAGGAACTGGCTTCTCTGGACCC 3645
Qy 281 CysTrpArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
Db TGTACCAGCAGCAGCAGTATCAAGACATCTCGCGATAACAACAACAGTGAATACTCG 3705
Qy 301 TrpThrGlyAlaSerLysTrpAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThr 320
Db TGGACTGGAGCTACCAAGTACCACTCAATGCGAGAGACTCTCTGGTGAATCCGGGGCCC 3765
Qy 321 AlaMetAlaSerHisLysAspAspGluAspLysPhePheProMetSerGlyValMetIle 340
Db GCCATGGCAGCCACAGGACGATGAAGAAATTTTCTCTCAGAGCGGGTTCTCATC 3825
Qy 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
Db TTTGGGAAGCAAGGCTCAGAGAAACAATGTGAACATTGAAAGGTCTATGATTACAGAC 3885
Qy 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
Db GAAGAGGAATTCGGAACCAACCAATCCCGTGGCTACGGAGCAGATATGGTTCTGTATCTACC 3945
Qy 381 AsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
Db AACCTCCAGAGAGCAACAGCAACGACGATACCGCAGATGTCAACACACACAGCGTCTTT 4005
Qy 401 ProGlyMetValTrpGlnAspArgAspValTrpLeuGlnGlyProIleTrpAlaLysIle 420
Db CCAGGATGCTGTCGACGACAGAGATGTGTACTTCAAGGGGCCATCTGGGCAAGATT 4065
Qy 421 ProHisLeuAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
Db CCACACGACGACGACATTTTCCCTCTCCCTCATGGGTGGATTCGACTTAACAC 4125
Qy 441 ProProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460
Db CCTCTCTCCACAGATTTCTCATCAAGAAACACCCCGTACTCTGCGAATCTCTCGACCACTTC 4185
Qy 461 SerAlaThrLysPheAlaSerPheIleThrGlnTrpSerThrGlyGln-ValSerValGln 480
Db AGTGGCGCAAGATTTGCTTCTTCATCACACAGTACTCCACGGGACACGCTCAGCGTGA 4245
Qy 480 uIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTrpTh 500
Db GATCGAGTGGAGCTGCAGAGGAAACAGCAACGCTGGATCCCGAATCCGGAATTCAGTACAC 4305
Qy 500 rSerAsnTrpAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTrpTh 520
Db TTCCAACACTACAAGTCTGTTAATCGTGGACTTACCGTGGATACTAATGGCGGTGATTTC 4365
Qy 520 rGluProArgProIleGlyThrArgTrpLeuThrArgProLeu 534
Db AGAGCGCTCGCCCATTTGGCAGCAGATACCTGATCTCGTAATCTG 4408

```

RESULT 8

```

; Sequence 1, Application PC/TUS9507178
; GENERAL INFORMATION:
; APPLICANT: Johnson, Philip R.
; TITLE OF INVENTION: Adeno-Associated Virus Materials and
; TITLE OF INVENTION: Methods
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago

```

STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07178

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Noland, Greta E.

REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 31975

TELECOMMUNICATION INFORMATION:

TELEPHONE: (312) 474-6300

TELEFAX: (312) 474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4680 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

PCT-US95-07178-1

Alignment Scores:

Pred. No.: 1.63e-266 Length: 4680
Score: 2439.50 Matches: 441
Percent Similarity: 89.16% Conservative: 36
Best Local Similarity: 82.43% Mismatches: 56
Query Match: 83.95% Indels: 2
DB: 5 Gaps: 1

US-09-807-802A-17 (1-534) x PCT-US95-07178-1 (1-4680)

QY 1 MetAlaSerGlyGlyAlaProMetAlaAspAsnGluGlyAlaValGly 20
DB 2809 ATGGCTACAGCGAGTGGCGCACCACCAATGGCAGACAAATACGAGGCGCGCGAGTGGGT 2868
QY 21 AsnAlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThr 40
DB 2869 AATTCCTCCGAAATGGCAATTCGATTCACATGGATGGCGCAGACAGTCAATCACCACC 2928
QY 41 SerThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSer 60
DB 2929 AGCACCCGAAACCTGGGCGCTGGCCACCTACACACACCACTCTACAAACAAATTTCCAGC 2988
QY 61 AlaSerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyr 80
DB 2989 CAATCA--CGAGCTCGAAGCACAATCACTACTTTGGCTACAGCACCCTTCGGGGGTAT 3045
QY 81 PheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsn 100
DB 3046 TTTGACTTCAACAGATTCCACTCGCCACTTTTCCACCGCTGACTGGCAAGACTCATCAAC 3105
QY 101 AsnAsnTrpGlyPheArgProLysArgLeuAsnPhelyleuPheAsnIleGlnValLys 120
DB 3106 AACAACTGGGGATTCGACCCCAAGAGACTCACTTCAAGCTCTTTAATCAATTCAGTCMAA 3165
QY 121 GluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeuThrSerThrValGln 140
DB 3166 GAGGTACGCGAATGACGCTACGACGACGATTCGCCAATAACCTTACCAGCAGCGTTCCAG 3225
QY 141 ValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCys 160
DB 3226 GTGTTTACTGACTCGGAGTACCAGCTCCCGTACGCTCCGCGTCCGCGCATCAAGATGC 3285
QY 161 LeuProProPheProAlaAspValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsn 180

DB 3286 CTCGCGCGCTTCCCGCAGCAGCGTCTTCTCATGGTGCCACAGTATGGATACCTCACTCCCTGAAC 3345
QY 181 AsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPheProSerGln 200
DB 3346 AACGGGAGTCCAGGAGTAGGAGCGCTCTTCACTTTACTGCTGGAGTACTTTCTTCTCTCAG 3405
QY 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValProPheHis 220
DB 3406 ATGCTGCGTACCGGAAACACACTTTTACCTTACGTACACTTTTGGAGGAGCTCTCTTCCAC 3465
QY 221 SerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyr 240
DB 3466 AGCAGTACGCTCAGCGCAGAGCTGACCGCTCTCATGATCTCTCTCATGACACGATAC 3525
QY 241 LeuTyrTyrLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeu 260
DB 3526 CTGTATTACTTTGAGCAGAACCAACACTCCAAGTGGACCAACCGCAGCTCAAGGCTTCAG 3585
QY 261 PheSerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyPro 280
DB 3586 TTTTCTCAGGCGGAGCGAGTGACATTCGGACCACTTAGGAACTGCTCTCTCTGAGCCC 3645
QY 281 CysTyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThr 300
DB 3646 TGTTACCCCGCAGCAGCGATTCAAAGACATCTCGGGATACACACACAGTGAATACTCG 3705
QY 301 TrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIleAlaAsnProGlyThr 320
DB 3706 TGGACTGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTGGTGAATCCGGGGCCC 3765
QY 321 AlaMetAlaSerHisLysAspAspClnAspLysPhePheProMetSerGlyValMetIle 340
DB 3766 GCCATGGCAAGCCACAGGAGCATGAGAAAGATTTTCTCTCAGAGCGGGGTCTCTATC 3825
QY 341 PheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAsp 360
DB 3826 TTTGGAGACGAGGCTCAGAGAAACAAATGTGAACATTTGAAAGGTCTATGATTACAGAC 3885
QY 361 GluGluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaVal 380
DB 3886 GAAGAGGAAATCGGAACAACCAATCCCGTGGCTAGCGAGCAGTATGTTCTCTATCTACC 3945
QY 381 AsnPheGlnSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeu 400
DB 3946 AACCTCCAGAGAGCGCAACAGACAGCAGCTACCGCAGATGTCAACACACACAGCGCTCT 4005
QY 401 ProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrpAlaLysIle 420
DB 4006 CCAGGCATGGTCTGCGAGGACAGAGATGTGTACTTTCAGGGGCCCATCTGGGCAAGATT 4065
QY 421 ProHisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsn 440
DB 4066 CCACACAGGACGAGCAATTTTCCACCCCTCTCCCTCATGGGTGGATTCGAGCTTAAACAC 4125
QY 441 ProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPhe 460
DB 4126 CCTCTCCACAGATTTCTCATCAAGAACACCCCGGTACTTCGGAATCTTTCGACCACTTC 4185
QY 461 SerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGln-ValSerValG 480
DB 4186 AGTGGGCAGAAATTTGCTTCTTCATCACACAGTACTCCACGGGACACCGTCCAGCGTGA 4245
QY 480 uIleGluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyrTh 500
DB 4246 GATCAGTGGGAGCTGCAGAGGAAAAACAGCAACGCTGGAATCCGGAATTCGGAATTCAGTAC 4305
QY 500 rSerAsnTyrAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyrTh 520
DB 4306 TTTCAACTACAAACAAGTCTGTTAAATCGTGGAGCTTACCGGTGGATCTAATATGGCGTGTATTC 4365
QY 520 rGluProArgProIleGlyThrArgTyrLeuThrArgProLeu 534
DB 4366 AGAGCTCGCCCATTTGGCACCAGCATCTGACTCGTAAATCTG 4408

RESULT 9
US-09-438-268-2
; Sequence 2, Application US/09438268
; Patent No. 6491907
; GENERAL INFORMATION:
; APPLICANT: Rabinowitz, Joseph B.
; APPLICANT: Samulski, Richard J.
; APPLICANT: Xiao, Weidong
; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING
; FILE REFERENCE: 5470-186
; CURRENT APPLICATION NUMBER: US/09/438,268
; EARLIER FILING DATE: 1999-11-10
; EARLIER FILING DATE: 1998-11-10
; EARLIER FILING DATE: 1999-03-10
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 8151
; TYPE: DNA
; ORGANISM: Virus
US-09-438-268-2

Alignment Scores:
Pred. No.: 1,59e-245 Length: 8151
Score: 2258.50
Percent Similarity: 85.18%
Best Local Similarity: 75.98%
Query Match: 77.72%
DB: 4
Gaps: 2

US-09-807-802A-17 (1-534) x US-09-438-268-2 (1-8151)

QY 2 AlaSerGlyGlyGlyAlaProMetAlaAspAsnGluGlyAlaAspGlyValGlyAsn 21
DB 2730 GCAGCAGCTGGCGAGCTGCAGTC---GAGGCGGACAGAGTCCCATGGAGTGGGTAT 2786
QY 22 AlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrSer 41
DB 2787 GCCTCGGGTATGGCATTCGATTCGATTCACCTGCTGAGGCGCCACGTCACGACAC 2846
QY 42 ThrArgThrTrpAlaLeuProThrTyraAsnHisLeuTyrylsGlnIleSerAla 61
DB 2847 ACCAGAACCTGGGTCTTGGCCACCTCAACCAACACCTCTACAGGACCTCGAGAG--- 2903
QY 62 SerThrGlyAlaSerAsnAspAsnHisTyryPheGlyTyzSerThrProTrpGlyTyPhe 81
DB 2904 -----AGCCTGCAGTCCACACCTACACGAGTCTCCACCCCTGGGATCTTT 2954
QY 82 AspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsnAsn 101
DB 2955 GACTTCAACCGCTTCCATCCGCTTCTCACCAGCTGACTGGCAGGAGCTCATCAACAAC 3014
QY 102 AsnTrpGlyPheArgProTyraArgLeuAsnPheIysLeuPheAsnIleGlnValLysGlu 121
DB 3015 AACTGGGGCATGGCACCAAGCCATCGGGTCAAAATCTTCAACATCCAGGTCAGGAG 3074
QY 122 ValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGlnVal 141
DB 3075 GTCAAGCAGCTCGAAGCGGACACACCGTGGCTTAATAACCTTACCAGCAGGTTCCAGATC 3134
QY 142 PheSerAspSerGluTyryGlnLeuProTyryrValLeuGlySerAlaHisGlnGlyCysLeu 161
DB 3135 TTTGGCAGCTCGTGTACGAACCTGCCGTACGTCTCCGCTCGGCGCATCAAGATGCCCTC 3194
QY 162 ProProPheProAlaAspValPheMetIleProGlnTyryGlyTyryrLeuThrLeuAsnAsn 181
DB 3195 CCGCGGTCACGACGAGCTCTCATGGTGCACAGTATGATACCTCACCTCCATCAACAC 3254
QY 182 GlySerGlnAlaValGlyArgSerSerPheTyryCysLeuGluTyryPheProSerGlnMet 201

Db 3255 GCGAGTCAGGACGAGTACCTCTTCATTTTACCTGCTGGAGTACTTCTCTCAGATG 3314
QY 202 LeuArgThrGlyAsnAsnPheThrPheSerTyryrThrPheGluGluValProPheHisSer 221
DB 3315 CTGCGTACCGGAAACAACATTTTACCTTACCTTACCTTGTGAGGAGCTTCTCTTCCACAGC 3374
QY 222 SerTyryAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyryLeu 241
DB 3375 AGCTACGCTACAGCAGGAGTGGACCTCTCATGAATCTCTCATGACCGAGTACCTG 3434
QY 242 TyryTyryLeuAsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLeuAspLeuPhe 261
DB 3435 TATTACTTGGACGACAGAACACACCTCAAGTGGACACACCGACAGCTCAAGGCTTCA 3494
QY 262 SerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyProCys 281
DB 3495 TCTCAGCGCGGAGCGAGTGACATTCGGGACGAGTCTAGGAACCTGGCTTCTCTGGACCTGT 3554
QY 282 TyryArgGlnGlnArgValSerIysThrLysThrAspAsnAsnAsnSerAsnPheThrTrp 301
DB 3555 TACCGCCAGCAGCGAGTATCAAGACATCTCGCGGATAACAACACAGTGAATCTCTG 3614
QY 302 ThrGlyAlaSerLysTyryAsnLeuAsnGlyArgGluSerIleIleAsnProGlyThrAla 321
DB 3615 ACTGGAGCTACCAAGTACCACCTCAATGCGAGAGACTCTCTGGTGAATCCGGGCGCGGCC 3674
QY 322 MetAlaSerHisLysAspAspGluAspLysPhePheProMetSerSerGlyValMetIlePhe 341
DB 3675 ATGGCAAGCCACACAGGACGATGAAGAAAGTTTTTCTCTCAGAGCGGGTCTCTCATCTTT 3734
QY 342 GlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAspGlu 361
DB 3735 GGGAAAGCAAGGCTCAGAGAAACAATGTGAACATTTGAAGAGTCTCATGTATACACAGCNA 3794
QY 362 GluGluIleLysAlaThrAsnProValAlaThrGluArgPheGlyThrValAlaValAsn 381
DB 3795 GAGGAAATCGGAACAACCAATCCCTGCTACGAGGACGATGTGTCTGTATCTACCAAC 3854
QY 382 PheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeuPro 401
DB 3855 CTCCAGAGAGGCAACAGACAGCAGTACCGCAGATGTCAACACACAGGCGCTTCTTCCA 3914
QY 402 GlyMetValTrpGlnAspArgAspValTyryLeuGlnGlyProIleTrpAlaLysIlePro 421
DB 3915 GGCATGCTTGGCAGCAGAGATGTGTACCTTACAGGGGCCCATCTGGGCAAGATTCCA 3974
QY 422 HisThrAspGlyHisPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsnPro 441
DB 3975 CACAGGACGAGCATTTTCCACCCCTCTCCCTCATGGGTGGATTCGGACTTAACACCTT 4034
QY 442 ProProGlnIleLeuIleLysAsnThrProValProAlaAsnProProAlaGluPheSer 461
DB 4035 CCTCACAGATTTCTCATCAAGAACACCCCGGTACTTGGATCTTCGACCTTCAGTTCAGT 4094
QY 462 AlaThrLysPheAlaSerPheIleThrGlnTyryrSerThrGlyGlnValSerValGluIle 481
DB 4095 GCGGCAAGTTTGTCTTCTTCATCACAGTACTCCAGGGACAGGTTCAGCTGGAGATC 4154
QY 482 GluTrpGluLeuGlnLysGluAsnSerLysArgTrpAsnProGluValGlnTyryrThrSer 501
DB 4155 GAGTGGGAGCTCAGAGGAAACAGCAACCGCTGGAATCCGGAATTCAGTACACTTCC 4214
QY 502 AsnTyryAlaLysSerAlaAsnValAspPheThrValAspAsnAsnGlyLeuTyryrThrGlu 521
DB 4215 AACTACACAGTCTGTATCTGTGACTTACCGTGGATCTATAGCGGTGTATTACAGAG 4274
QY 522 ProArgProIleGlyThrArgTyryrLeuThrArgProLeu 534
DB 4275 CCTCGCCCCATTGGCACCAGATACCTGACTCGTAAATCTG 4313
RESULT 10
US-09-532-594B-19

```

; Sequence 19, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.
; APPLICANT: Kotin, Robert M.
; APPLICANT: Safer, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.0252U2
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 1617
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; OTHER INFORMATION: synthetic construct
; NAME/KEY: misc_feature
; LOCATION: 162
; OTHER INFORMATION: n = g, a, c or t(u)
; NAME/KEY: misc_feature
; OTHER INFORMATION: AAV4 capsid protein VP3 gene
; US-09-532-594B-19

Alignment Scores:
Pred. No.: 4,198-182 Length: 1617
Score: 1690.50 Matches: 325
Percent Similarity: 71.30% Conservative: 65
Best Local Similarity: 59.41% Mismatches: 132
Query Match: 58.17% Indels: 25
DB: 4 Gaps: 9

US-09-807-802A-17 (1-534) x US-09-532-594B-19 (1-1617)

QY 2 AlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGlyAsn 21
Db 7 GCAGCAGCTGGCGAGTGCAGTC--GAGGSGGACAAAGTGGCGATGGGTAAAT 63
QY 22 AlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrSer 41
Db 64 GCCTCGGGTGATGGCATTTGCCATTTCCACCTGGTCTGAGGGGCCACGTCACGACCACG 123
QY 42 ThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerAla 61
Db 124 ACCAGAACCTGGTCTTGCCCACTTACAAACACCCTTACAGGCTCGGAGAG-- 180
QY 62 SerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyrPhe 81
Db 181 -----AGCCTGCAGTCCAAACACCTACACGGATTTCTCACCCCTGGGGTACTTT 231
QY 82 AspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsnAsn 101
Db 232 GACTTCAACCGTTCCATGCCACCTCTCACCGTGAAGTGGCGGCTCATCAACCAAC 291
QY 102 AsnTrpGlyPheArgProLysArgLeuAsnPheLeuPheAsnIleGlnValLysGlu 121
Db 292 AACTGGGGCATCGGACCCAAAGCCATGCGGGTCAAAATCTTCAATCATCGAGTCAAGGAG 351
QY 122 ValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGlnVal 141
Db 352 GTCAAGCAGCTCGAACCGGAGACAAAGTGGGTAAATTAACCTTACCAGCAGCGTTCCAGATC 411
QY 142 PheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCysLeu 161
Db 412 TTTGGGACTCTGCTGACAACTGGCGTACGTGATGGATGGCGGTCAAGAGGCGACCTG 471
QY 162 ProProPheProAlaAspValPheMetIleProGlnTyrGlyTyr-----LeuThr 178
Db 472 CTCTCTTTTCCCAACGACGCTTTATGGTGCCCAAGTACGGGTACTGTGGACTGTGTGACC 531

```

```

QY 179 LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 198
Db 532 GGCAACACTTCGCAGCAACAGACTGACAGAAATGCTTCTACTGCTGGAGTACTTTCT 591
QY 199 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 218
Db 592 TCGCAGATGCTGGGACTGGCAACAACCTTTGAAATAGCTACAGTTTGGAGAGGTGCT 651
QY 219 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 238
Db 652 TTCCACTCGATGATACGGCAGCAGCAGCTGACCGGCTGATGAAACCTTCTCATCAG 711
QY 239 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsn-----GlnSerGlySerAla 254
Db 712 CAGTACCTGTGGGACTGCAATCGACACCAACCGGACCACTCTCAATAGTGGGACTGCC 771
QY 255 GlnAsnLysAspLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLys 274
Db 772 ACCACCAAC-----TTTACCAAGCTGGGCTTACCAACTTTTCAACTTTTAAAG 822
QY 275 AsnTrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsn 294
Db 823 AACTGGCTGCCCGGCTTCAATCAAGCAGCAGGGCTTCTCAAGACTGCC-----AAT 876
QY 295 AsnAsnSerAsnPheThrTrpThrGlyAlaSer-----LysTyr----- 307
Db 877 CAAAACCTACAAGATCCCTGCCACCGGGTCAAGACTCTCATCAATAGCAGACGACGAC 936
QY 308 AsnLeuAsnGlyArgGluSerIleAsnProGlyThrAlaMetAlaSerHisLysAsp 327
Db 937 ACTCTGGACGGAAGATGGAGTGGCTGACCCCGGACCTCCAAATGGCCACCGCTGGAG 996
QY 328 AspGluAspLysPhePheProMetSerGlyValMetIlePheGlyLysGluSerAlaGly 347
Db 997 CGGACACGACAG--TTACGCAACAGCAGCTCATCTTGGCGGGCTTAAACAGACGCG 1053
QY 348 AlaSerAsnThrAlaLeuAspAsnValMetIleThrAspGluGluGluIleLysAlaThr 367
Db 1054 AACACGGCCACCTAGTCCCGGACTCTGATCTTCACCTCTGAGGAGGAGCTGGCAGCAC 1113
QY 368 AsnProValAlaThrGluArgPheGlyThrValAlaValAsnPheGlnSerSerThr 387
Db 1114 AAGCCACCGATACGAGATGTGGGGCACTTCTGGCGGTGACCAAGCAAGCAAC 1173
QY 388 AspProAlaThrGlyAspValHisAlaMetGlyAlaLeuProGlyMetValTrpGlnAsp 407
Db 1174 CTGCGGACCGTGGACAGACTGACAGCTTGGGAGCGGCTGGAAATGCTCTGGCAAAAC 1233
QY 408 ArgAspValTyrLeuGlnGlyProIleTrpAlaLysIleProHisThrAspGlyHisPhe 427
Db 1234 AGAGACATTTACTACAGGGTCCCAATTTGGGCGCAAGATTCTCATACCGATGGACACTTT 1293
QY 428 HisProSerProLeuMetGlyPheGlyLeuLysAsnProProGlnIleLeuIle 447
Db 1294 CACCCCTCACCGCTGATTTGGTGGTGGCTGAAACACCCGCTCTCTCAATTTTATC 1353
QY 448 LysAsnThrProValProAlaAsnProProAlaGluPheSerAlaThrLysPheAlaSer 467
Db 1354 AAGAACACCCCGTACCTGCGAATCTCTCAACGACCTTTCAGCTCTCTCCGTAACCTCC 1413
QY 468 PheIleThrGlnTyrSerThrGlyGlnValSerValGluIleGluTrpLeuLeuLys 487
Db 1414 TTCATTACTAGTACAGCACTGGCCAGGTGCGGTGCAAGTTGATGGAGATCCAGAG 1473
QY 488 GluAsnSerLysArgTrpAsnProGluValGlnTyrThrSerAsnTyrAlaLysSerAla 507
Db 1474 GAGCGTCCAAACCGCTGGAACCCGAGGTCCAGTTTACTCTCAACTACGACGACAAAC 1533
QY 508 AsnValAspPheThrValAspAsnAsnGlyLeuTyrThrGluProArgProIleGlyThr 527
Db 1534 TCTCTGTTGTGGCTCCCGATCGCGCTGGGAAATACACTGAGCTAGGCTATCGGTACC 1593
QY 528 ArgTyrLeuThrArgProLeu 534

```

```
Db 1594 CGTACTCTCACCACCACTTG 1614
RESULT 11
US-09-532-594B-17
; Sequence 17, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.
; APPLICANT: Kottin, Robert M.
; APPLICANT: Safer, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.025202
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1800
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; OTHER INFORMATION: synthetic construct
; NAME/KEY: misc_feature
; LOCATION: 342
; OTHER INFORMATION: n = g, a, c or t(u)
; NAME/KEY: misc_feature
; OTHER INFORMATION: AAV4 capsid protein VP2 gene
US-09-532-594B-17

Alignment Scores:
Pred. No.: 5,02e-182 Length: 1800
Score: 1690.50 Matches: 325
Percent Similarity: 71.30% Conservative: 65
Best Local Similarity: 59.41% Mismatches: 132
Query Match: 58.17% Indels: 25
DB: 4 Gaps: 9

US-09-807-802A-17 (1-534) x US-09-532-594B-17 (1-1800)
Qy 2 AlaSerGlyGlyAlaProMetAlaAspAsnGluGlyAlaAspGlyValGlyAsn 21
Db 187 GCAGCAGCTGGCGAGGTGAGTGC--GAGGGGGACAAAGTGGCGATGAGTGGTAA 243
Qy 22 AlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrSer 41
Db 244 GCCTGGGGTATTGGCATTTCCATTCCACCTGCTGAGGGGCCACGTCACGACCCAGC 303
Qy 42 ThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSerAla 61
Db 304 ACCAGAACCTGGGTCTTGGCCACCCTACAAACACCCCTNTACAGGCAGCTCGGAGAG-- 360
Qy 62 SerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyrPhe 81
Db 361 -----AGCTGCGAGTCCAAACACCTACAAACGATTCTCCACCCCTCGGGATCTTT 411
Qy 82 AspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsnAsn 101
Db 412 GACTTCAACCGCTTCCACTGCGCACTTCTCACACGCTGACTGGCGACGACTCATCAACAAC 471
Qy 102 AsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLysGlu 121
Db 472 AACTGGGGCATGCCACCCAAAGCCATGGGGGTCAAAATCTTCAACATCCAGGTCAAGAG 531
Qy 122 ValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGlnVal 141
Db 532 GTCACGACGTCGAACGGCGAGACAAACGCTGCTTAATAACCTTACCAGCAGGTTCCAGATC 591
Qy 142 PheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCysLeu 161
Db 592 TTTGGGACTCGTGTACGAACCTGCGGTACGTGTGATGGGTGCGGGTCAAGAGGGGAGCCTG 651
Qy 162 ProProPheProAlaAspValPheMetIleProGlnTyrGlyTyr-----LeuThr 178
Db 652 CCTCTTTTCCCAAGCAGCTCTTTATGTGTCCTCCAGTACGGCTACTGTGGACTGGTGACC 711
Qy 179 LeuAsnAsnGlySerGlnAlaValGlyArgSerPheTyrCysLeuGluTyrPhePro 198
Db 712 GGCAACACTTCGCAGCAACAGACTGACAGAAATGCTTCTACTGCTCGAGTACTTCTCT 771
Qy 199 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 218
Db 772 TCGCAGATGCTGCGGACTGGCAACACTTTGAAATACGTACAGTTTGTGAGAAGGTGCT 831
Qy 219 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 238
Db 832 TTCCACTCGATGTAGCGCACAGCAGCTGAGCCGCTGATGAAACCTCTCATCGAC 891
Qy 239 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsn-----GlnSerGlySerAla 254
Db 892 CAGTACCTGTGGGACTGCAATCGACCACCCGGAACCCCTGAATGCGGGAGCTGCC 951
Qy 255 GlnAsnLysAspLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLys 274
Db 952 ACCCAAC-----TTTACCAGCTGGGCTTACCAACTTTTCAACTTTTAAAG 1002
Qy 275 AsnTrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsn 294
Db 1003 AACTGGCTGCGCGGCTTCAATCAAGCAGCAGGGCTTCTCAAGACTGCC-----AAT 1056
Qy 295 AsnAsnSerAsnPheThrTrpThrGlyAlaSer-----LysTyr----- 307
Db 1057 CAAAACCTACAGATCCCTGCCACCGGTCAGACAGTCTCATCAATACGAGACGCACAGC 1116
Qy 308 AsnLeuAsnGlyArgGluSerIleLeuAsnProGlyThrAlaMetAlaSerHisLysAsp 327
Db 1117 ACTCTGGAGCGAAGATGGAGTGCCTGACCCCGGACCTCCAAATGGCCAGCGCTGACCT 1176
Qy 328 AspGluAspLysPhePheProMetSerGlyValMetIlePheGlyLysGluSerAlaGly 347
Db 1177 GCGGACAGCAAG--TTCAGCAACAGCCAGCTCATCTTTGCGGGGCTTAAACAGAACGCG 1233
Qy 348 AlaSerAsnThrAlaLeuAspAsnValMetIleThrAspGluGluLeuLysAlaThr 367
Db 1234 AACACGGCCACCGTACCGGGACTCTGATCTTTCACCTCTGAGGAGGAGCTGCGACGCC 1293
Qy 368 AsnProValAlaThrGluArgPheGlyThrValAlaValAsnPheGlnSerSerThr 387
Db 1294 AACGCCACCGATACGAGCATGTGGGGCAACCTTACCTGGCGGTGACCAAGCAACGCAAC 1353
Qy 388 AspProAlaThrGlyAspValHisAlaMetGlyAlaLeuProGlyMetValTrpGlnAsp 407
Db 1354 CTGCGGACCGTGCAGACAGCTGACAGCTTGGGAGCGCTGCTGGAATGCTCTGGCAAAAC 1413
Qy 408 ArgAspValTyrLeuGlnGlyProIleTrpAlaLysIleProHisThrAspGlyHisPhe 427
Db 1414 AGAGACATTTACTACCGAGGTCTCCATTTGGGGCAAGATTCCTCATACCGATGACACTTT 1473
Qy 428 HisProSerProLeuMetGlyGlyPheGlyLeuLysAsnProProGlnIleLeuIle 447
Db 1474 CACCCCTCACCGCTGATTGGTGGTGGCTGGAACACCCGCTCTCAATTTTATC 1533
Qy 448 LysAsnThrProValProAlaAsnProProAlaGluPheSerAlaThrLysPheAlaSer 467
Db 1534 AAGAACACCCCGTACTCGAATCTCGCAACGACCTTCAGCTCTACTCCGGTAAACTCC 1593
Qy 468 PheIleThrGlnTyrSerThrGlnValSerValGluIleGluTrpGluLeuGlnLys 487
Db 1594 TTCAATTAATCAGTACAGCAGCTGCGCAGGTGTCGGTGCAGATTGATGAGGAGATCCAGAAG 1653
Qy 488 GluAsnSerLysArgTrpAsnProGluValGlnTyrThrSerAsnTyrAlaLysSerAla 507
Db 1654 GAGCGGTCCAAACGCTGGAAACCCCGAGGTCCAGTTTACCTCCACTACGACAGCAAAAC 1713
```

| | | | |
|----|------|--|------|
| Qy | 508 | AsnValAspPheThrValAspAsnAsnGlyLeuTyrThrGluProArgProIleGlyThr | 527 |
| | | | |
| | | | |
| | | | |
| | | | |
| Db | 1714 | TCCTCTGTGTGGGCTCCCGATGCGGCTGGGAAATACACTGAGCCTAGGGCTATCGGTACC | 1773 |
| | | | |
| | | | |
| | | | |
| Qy | 528 | ArgTyrLeuThrArgProLeu | 534 |
| | | | |
| Db | 1774 | CGTACCTCTACCCACCACCTG | 1794 |

RESULT 12

```

US-09-532-594B-5
; Sequence 5, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.
; APPLICANT: Kotin, Robert M.
; APPLICANT: Safer, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.0252U2
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 2208
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; OTHER INFORMATION: synthetic construct
; NAME/KEY: misc_feature
; LOCATION: 750
; OTHER INFORMATION: n = g, a, c or t(u)
; NAME/KEY: misc_feature
; OTHER INFORMATION: AAV4 capsid protein VP1 gene
US-09-532-594B-5

```

| | |
|------------------------|-----------|
| Alignment Scores: | |
| Pred. No.: | 7,07e-182 |
| Score: | 1690.50 |
| Percent Similarity: | 71.30% |
| Best Local Similarity: | 59.41% |
| Query Match: | 58.17% |
| DB: | 4 |
| | 9 |
| Length: | 2208 |
| Matches: | 325 |
| Conservative: | 65 |
| Mismatches: | 132 |
| Indels: | 25 |
| Gaps: | 9 |

```

US-09-807-802A-17 (1-534) x US-09-532-594B-5 (1-2208)

Qy      2 AlaSerGlyGlyValAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGlyAsn 21
Db          ::::: ::: :
S95 GCACAGCTGGCGGAGCTGCAGTC--GAGGSGCACAAGGTGCCGATGGAGTGGGTAAAT 651
Qy      22 AlaSerGlyAsnThrHisCyAspSerThrTrpLeuGlyAspArgValIleThrThrSer 41
Db          ::::: :::::
S62 GCCTCGGGTGATTGGCATTTCCACTGGTCTGAGGGCCACGTCAAGACCACACGAGC 711
Qy      42 ThrArgThrTrpAlaLeuProThrTyrrAsnAsnHisLeuTyrLysGlnIleSerSerAla 61
Db          ::::: :::::
W12 ACCAGAACCTGGGTCTTGCCCACTTACAACAACCACTNTTACAAGCGACTCGGAGAG--- 768
Qy      62 SerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyrPhe 81
Db          :::::
V69 -----AGCGTCGACTCCAACACCTTACAACGGGATTTCTCCACCCCCTGGGGATACATT 819
Qy      82 AspPheAsnArgPheHisCyHisPheSerProArgAspTrpGlnArgLeuIleAsnAsn 101
Db          ::::: :::::
S20 GACTTCAAACCCGCTTCCATCGCCACTTCTCACCACTGACTGGCAGCGACTCATCAACAAC 879
Qy      102 AsnTrpGlyPheArgProLyArgLeuAsnPheLysLeuPheAsnIleGlnVallysGlu 121
Db          ::::: :::::
S80 AACTGGGGCATCGGNACCCGAAGCCATGCGGGTCAAAAATCTTCAACATCCAGGTCAAGGAG 939
Qy      122 ValThrThrAsnAspGlyValThrThrThrIleAlaAsnAsnLeuThrSerThrValGlnVal 141
Db          ::::: :::::

```

```

QY 488 GluAenSerLysArgTrpAenProGluValGlnTyrThrSerAenTyrAlaLysSerAla 507
DB 2062 GAGCGGTCCAAACGCTGGAAACCCCGAGGTCAGTTTACCTCCAACTACGACAGCAAAAC 2121
QY 508 AenValAspPheThrValAlaAspAenAenGlyLeuTyrThrGluProArgProIleGlyThr 527
DB 2122 TCTCTGTTGGGCTCCCATGCGGTGGGAATAACACTGAGCTAGGGCTATCGGTACC 2181
QY 528 ArgTyrLeuThrArgProLeu 534
DB 2182 CGTACCTCACCACCACCTG 2202

RESULT 13
US-09-532-594B-1
; Sequence 1, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.
; APPLICANT: Kotin, Robert M.
; APPLICANT: Safer, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.025202
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4767
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; OTHER INFORMATION: synthetic construct
; NAME/KEY: misc feature
; LOCATION: 3009
; OTHER INFORMATION: n = g, a, c or t(u)
; NAME/KEY: misc feature
; OTHER INFORMATION: AAV4 genome
US-09-532-594B-1

Alignment Scores:
Pred. No.: 2 576-181 Length: 4767
Score: 1690.50 Matches: 325
Percent Similarity: 71.30% Conservative: 65
Best Local Similarity: 59.41% Mismatches: 132
Query Match: 58.17% Indels: 25
DB: 4 Gaps: 9

US-09-807-802A-17 (1-534) x US-09-532-594B-1 (1-4767)
QY 2 AlaSerGlyGlyAlaProMetAlaAspAenAenGluGlyAlaAspGlyValGlyAsn 21
DB 2854 GCAGCAGCTGGCGAGCTGCAGTC--GAGGSGGACNAGGTCCGATGGAGTGGGTAAT 2910
QY 22 AlaSerGlyAenTyrHisCysAspSerThrTyrLeuGlyAspArgValIleThrThrSer 41
DB 2911 GCCTCGGGTATGGCATTGGGATTCACCTGGTCTGGGGGCCACGTCACGACCCAGC 2970
QY 42 ThrArgThrTrpAlaLeuProThrTyrAenAenHisLeuTyrLysGlnIleSerSerAla 61
DB 2971 ACCAGAACCTGGGTCTGGCCACCCTCAACAACACCCTNTACAAGCGACTCGGAGAG-- 3027
QY 62 SerThrGlyAlaSerAenAenAenHisTyrPheGlyTyrSerThrProTrpGlyTyrPhe 81
DB 3028 -----AGCCTGCAGTCCAAACACCTACAACGGATTCTCCACCCCTCGGGGATACTTT 3078
QY 82 AspPheAenArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAenAen 101
DB 3079 GACTTCAACCGCTTCCACTGCTCTCCACGAGTGGGAGCGGACTCATCAACAAC 3138
QY 102 AsnTrpGlyPheArgProLysArgLeuAenPheLysLeuPheAenIleGlnValLysGlu 121

```

```

DB 3139 AACTGGGGCATGCGACCAAGCCATCGGGGTCAAAATCTTCAACATCCAGCTCAAGGAG 3198
QY 122 ValThrThrAenAspGlyValThrThrIleAlaAenAenLeuThrSerThrValGlnVal 141
DB 3199 GTCAAGAGCTCGAAGCGCGAGACAACGGTGGCTAATAACCTTACGACGACCGTTTCAATC 3258
QY 142 PheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCysLeu 161
DB 3259 TTTGGGACTCGTGTAGTAACTGCTGCTGATGATGGATGGGTCAAGAGGGGAGCCCTG 3318
QY 162 ProPheProAlaAspValPheMetIleProGlnTyrGlyTyr-----LeuThr 178
DB 3319 CCTCTTTTCCCAACGAGCTCTTTATGTGCTCCCGAGTACGGCTACTGTGGACTGTGTGACC 3378
QY 179 LeuAenAenGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 198
DB 3379 GGCAACACTTCGCGACCAACAGACTGCACAGAAATGCCCTTCTACTGCTGGAGTACTTCTCT 3438
QY 199 SerGlnMetLeuArgThrGlyAenAenPheThrPheSerTyrThrPheGluGluValPro 218
DB 3439 TCGCAGATGCTGGCGACTGGCAACAACCTTGTAAATTAGGTACAGTTTGTAGAAGGTGCT 3498
QY 219 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAenProLeuIleAsp 238
DB 3499 TTCCACTCGATGTACGGCACAGCCAGAGCTGGACCGGCTGATGAACCCCTCTCATCGAC 3558
QY 239 GlnTyrLeuTyrTyrLeuAenArgThrGlnAen-----GlnSerGlySerAla 254
DB 3559 CAGTACCTGTGGGAGCTGCAATTCACACCAACCGGACCCCTGATGTCGGGAGCTGCC 3618
QY 255 GlnAenLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLys 274
DB 3619 ACCACCAAC-----TTTACCAGCTGGCGCTACCAACTTTTCCAACCTTTAAAAAG 3669
QY 275 AenTrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAen 294
DB 3670 AACTGGCTGCCCGGCTTCAATCAAGCAGCAGGGCTTCTCAAGAGACTGCC-----AAT 3723
QY 295 AenAenSerAenPheThrTyrThrGlyAlaSer-----LysTyr----- 307
DB 3724 CAAAACTACAAGATCCCTGCCCGGCTGCAGCAGCTCTCATCAATACGACGACGACGAC 3783
QY 308 AenLeuAenGlyArgGluSerIleAenProGlyThrAlaMetAlaSerHisLysAsp 327
DB 3784 ACTCTGGACGAGATGGAGTGCCTGACCCCGGACCTCCCAATGGCCACGCTGACCT 3843
QY 328 AspGluAspLysPhePheProMetSerGlyValMetIlePheGlyLysGluSerAlaGly 347
DB 3844 GCGGACAGCAAG---TTCAGCAACAGCCAGCTCATCTTTGGCGGCGCTTAAACAGAACGCGC 3900
QY 348 AlaSerAenThrAlaLeuAenAspAenValMetIleThrAspGluGluIleLysAlaThr 367
DB 3901 AACACGGCCACGTCATCCCGGAGCTCTGATCTTCACTCTGAGGAGAGACTGGCAGCCACC 3960
QY 368 AenProValAlaThrGluArgPheGlyThrValAlaValAenPheGlnSerSerThr 387
DB 3961 AACGCCACCGATACGACATGTGGGGCAACCTTACCTGGCGGTGACGACGACCAACGAC 4020
QY 388 AspProAlaThrGlyAspValHisAlaMetGlyAlaLeuProGlyMetValTrpGlnAsp 407
DB 4021 CTGCGCGCGGTGACAGACTGCACGCTTGGGAGCGCTGGGATGGTCTGGGCAAAAC 4080
QY 408 ArgAspValTyrLeuGlnGlyProIleTrpAlaLysIleProHisThrAspGlyHisPhe 427
DB 4081 AGAGACATTACTACAGGGTCCCAATTTGGGCAAGATTTCTCATCCGATGACGACACTTT 4140
QY 428 HisProSerProLeuMetGlyGlyPheGlyLeuLysAenProProGlnIleLeuIle 447
DB 4141 CACCCCTCACCCTGATTGGTGGGTTTGGGCTGAAACACCCCGCTCTCAAAATTTTATC 4200
QY 448 LysAenThrProValProAlaAenProProAlaGluPheSerAlaThrLysPheAlaSer 467

```

Db 4201 AAGAACACCCGGTACTCGGAATCTCTGCAACGACCTTCAGCTCTACTCGGTAACTCC 4260
Qy 468 PheileThrGlnTyrSerThrGlyGlnValSerValGluLeuIleGluTrpGluLeuGlnLys 487
Db 4261 TTCAATTACTCAGTACAGCACTGGCCAGGTGTCGGTGCAGATTGACCTGGGAGATCCAGAAG 4320
Qy 488 GluAsnSerLysArgTrpAsnProGluValGlnTyrThrSerAsnTyrAlaLysSerAla 507
Db 4321 GAGCGGTCCAAACGCTGGGAACCCCGAGGTCCAGTTTACCTCCAACTACCGACAGCAAAAC 4380
Qy 508 AsnValAspPheThrValAspAsnAsnGlyLeuTyrThrGluProArgProIleGlyThr 527
Db 4381 TCTCTGTTGGGCTCCCGATGGCGTGGGAATACATGAGCTAGGGCTATCGGTACC 4440
Qy 528 ArgTyrLeuThrArgProLeu 534
Db 4441 CGCTACCTCACCACCACCTG 4461

RESULT 14

US-09-438-268-1/c
; Sequence 1, Application US/09438268
; Patent No. 6491907
; GENERAL INFORMATION:
; APPLICANT: Rabinowitz, Joseph E.
; APPLICANT: Xiao, Weidong
; APPLICANT: Samulski, Richard J.
; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING
; FILE REFERENCE: 5470-186
; CURRENT APPLICATION NUMBER: US/09/438,268
; EARLIER APPLICATION NUMBER: 60/107,840
; EARLIER FILING DATE: 1998-11-10
; EARLIER APPLICATION NUMBER: 60/123,651
; EARLIER FILING DATE: 1999-03-10
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 7214
; TYPE: DNA
; ORGANISM: Virus
US-09-438-268-1

Alignment Scores:

Pred. No.: 5,16e-181 Length: 7214
Score: 1690.50 Matches: 325
Percent Similarity: 71.30% Conservative: 65
Best Local Similarity: 59.41% Mismatches: 132
Query Match: 58.17% Indels: 25
DB: 4 Gaps: 9

US-09-807-802A-17 (1-534) x US-09-438-268-1 (1-7214)

Qy 2 AlaSerGlyGlyGlyAlaProMetAlaAspAsnAsnGluGlyAlaAspGlyValGlyAsn 21
Db 2389 GCAGCAGCTGGCGAGCTGCAGTC---GAGGGCGGACAGGTCCCATGGAGTGGGTAAT 2333
Qy 22 AlaSerGlyAsnTrpHisCysAspSerThrTrpLeuGlyAspArgValIleThrThrSer 41
Db 2332 GCCTCGGGTGATTGGCATTTGGATTCCACCTGCTGAGGGCCACGTCACGACCACG 2273
Qy 42 ThrArgThrTrpAlaLeuProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSerAla 61
Db 2272 ACCAGAACCTGGGTCTTGCCACCTTACCAACACCTCTTACAGCGACTCGGAGAG--- 2216
Qy 62 SerThrGlyAlaSerAsnAspAsnHisTyrPheGlyTyrSerThrProTrpGlyTyrPhe 81
Db 2215 -----AGCTCGAGTCCCAACACTTACACGGATTCTTCAACCCCTCGGGGATACTTT 2165
Qy 82 AspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArgLeuIleAsnAsn 101
Db 2164 GACTTCAACCGCTTCCACTGCGCACTTCTCACCACGCTGACTGGCAGGACTCATCAACAAC 2105

Qy 102 AsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIleGlnValLysGlu 121
Db 2104 AACTGGGCGATCGCGACCCCAAGCCATCGCGGTCAAAATCTTCAACATCCAGGTCAAGGAG 2045
Qy 122 ValThrThrAsnAspGlyValThrThrIleAlaAsnAsnLeuThrSerThrValGlnVal 141
Db 2044 GTCAGCAGCTCGACGGCGGACACACGGTGGCTAATAACCTTACACGACCGGTTCAGATC 1985
Qy 142 PheSerAspSerGluTyrGlnLeuProTyrValLeuGlySerAlaHisGlnGlyCysLeu 161
Db 1984 TTTGGGAGCTCGTCTGATGAACTCGCTACGATGGGTGCGGTCAAGAGGGCGACCTG 1925
Qy 162 ProProPheProAlaAspValPheMetIleProGlnTyrGlyTyr-----LeuThr 178
Db 1924 CCTCTTTTCCCAACGACGCTTTATGTATGGCCAGTACGGCTACTGTGGAGTGTGTGACC 1865
Qy 179 LeuAsnAsnGlySerGlnAlaValGlyArgSerSerPheTyrCysLeuGluTyrPhePro 198
Db 1864 GGCAACACTTCGCAGCAACAGACTGCAGAAATGCTTCTACTGCTGGAGTACTTTCT 1805
Qy 199 SerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGluGluValPro 218
Db 1804 TCGCAGATGCTGCGGACTGGCAACAACCTTGAATATACGTACAGTTTTCAGAGAGTGCT 1745
Qy 219 PheHisSerSerTyrAlaHisSerGlnSerLeuAspArgLeuMetAsnProLeuIleAsp 238
Db 1744 TTCCACTCGATGTACGCGCACAGCAGCTGGACCGGCTGATGAACCTCTCATCGAC 1685
Qy 239 GlnTyrLeuTyrTyrLeuAsnArgThrGlnAsn-----GlnSerGlySerAla 254
Db 1684 CAGTACTGTGGGAGCTGCAATCGACACCCAGCAACCCCTCAATGCGGGAGTGC 1625
Qy 255 GlnAsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMetSerValGlnProLys 274
Db 1624 ACCACCAAC-----TTTACCAAGCTGGCGCTACCAACTTTTCAACTTTAAAAAG 1574
Qy 275 AsnTrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThrLysThrAspAsn 294
Db 1573 AACTGGCTGCGCGGCTTCAATCAAGCAGCAGGCGCTTCTCAAGACTGCCC-----AAT 1520
Qy 295 AsnAsnSerAsnPheThrTrpThrGlyAlaSer-----LysTyr----- 307
Db 1519 CAAAACTACAGATCCCTGCCACCGGTGTCAGACAGTCTCATCAATACGAGACGCACAGC 1460
Qy 308 AsnLeuAsnGlyArgGluSerIleIleAsnProGlyThrAlaMetAlaSerHisLysAsp 327
Db 1459 ACTCTGGACGGAAGATGGAGTGCCTGACCCCGGACCTCCCAATGGCCACGCTGGACCT 1400
Qy 328 AspGluAspLysPhePheProMetSerGlyValMetIlePheGlyLysGluSerAlaGly 347
Db 1399 GCGGACAGCAAG---TTCAGCAACAGCCAGCTCATCTTTGGCGGGCTTAAACAGAACGCG 1343
Qy 348 AlaSerAsnThrAlaLeuAspAsnValMetIleThrAspGluGluGluIleLysAlaThr 367
Db 1342 AACAGCGCCACCGTACCCCGGACTCTGATCTTCACTCTGAGGAGAGAGCTGCGACGCCAC 1283
Qy 368 AsnProValAlaThrGluArgPheGlyThrValAlaValAsnPheGlnSerSerSerThr 387
Db 1282 AACGCCACCGATACGACATGTGGGGCAACCTTACCTGGCGGTGACACAGCAACAGCAAC 1223
Qy 388 AspProAlaThrGlyAspValHisAlaMetGlyAlaLeuProGlyMetValTrpGlnAsp 407
Db 1222 CTGGCGACCGTGGACAGACTGACAGCTTGGGAGCGGCTGGAGTGGTCTGGCAAAAC 1163
Qy 408 ArgAspValTyrLeuGlnGlyProIleTrpAlaLysIleProHisThrAspGlyHisPhe 427
Db 1162 AGAGACATTTACTACCGAGGTCCCATTTGGGCCAAGATTTCTCATACCGATGACACTTT 1103
Qy 428 HisProSerProLeuMetGlyGlyPheGlyLeuLysAsnProProProGlnIleLeuIle 447
Db 1102 CACCCCTCAGCGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1043
Qy 448 LysAsnThrProValProAlaAsnProProAlaGluPheSerAlaThrLysPheAlaSer 467

```

Db      1042  AGAAACACCCCGTACTCGAATCTCGAACACCTTCAGCTCTACTCCGGTAAACTCC 983
Qy      PheileThrGlnTyrSerThrGlnValSerValGluilelelurpGluLeuGlnLys 487
Db      982  TTCATTACTCAGTACAGCACTGGCCAGGTGTGGTGCAGATTGACTGGGAGATCCAGAAG 923
Qy      488  GluAsnSerLysArgTrpAsnProGluValGlnTyrThrSerAsnTyrAlaLysSerAla 507
Db      922  GAGCGGTCCAAACGCTGGAAACCCCGAGGTCAGTTTACCTCCAACTACGGACAGCAAAAC 863
Qy      508  AsnValAspPheThrValAspAsnGlyLeuTyrThrGluProArgProileGlyThr 527
Db      862  TCCTGTGTGGCTCCGATCGGCTGGGAAATACATGAGCCTAGGCTATCGGTACC 803
Qy      528  ArgTyrLeuThrArgProLeu 534
Db      802  CGCTACTCACCACCACTG 782

RESULT 15
US-09-438-268-3
; Sequence 3, Application US/09438268
; Patent No. 6491907
; GENERAL INFORMATION:
; APPLICANT: Rabinowitz, Joseph B.
; APPLICANT: Samuleki, Richard J.
; APPLICANT: Xiao, Weidong
; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 5470-186
; CURRENT APPLICATION NUMBER: US/09/438,268
; CURRENT FILING DATE: 1999-11-10
; EARLIER APPLICATION NUMBER: 60/107,840
; EARLIER FILING DATE: 1998-11-10
; EARLIER APPLICATION NUMBER: 60/123,651
; EARLIER FILING DATE: 1999-03-10
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2271
; TYPE: DNA
; ORGANISM: Virus
US-09-438-268-3

Alignment Scores:
Pred. No.: 1,72e-45 Length: 2271
Score: 491.50 Matches: 144
Percent Similarity: 43.57% Conservative: 83
Best Local Similarity: 27.64% Mismatches: 239
Query Match: 16.91% Indels: 55
DB: 4 Gaps: 16

US-09-807-802A-17 (1-534) x US-09-438-268-3 (1-2271)
Qy      10  AlaAspAsnGluGlyAlaAspGlyValGlyAlaSer-----GlyAsnTrpHis 27
Db      625  GCAGAACCCAGCACTGTGTGAGGAGGGGGGGGAGTAACTCTGTCAAAACCACTGTGGAGT 684
Qy      28  CysAspSerThrTrpLeuGlyAspArgValileThrThrSerThrArgThrTrpAlaLeu 47
Db      685  GAGGGGGCCACTTTTAGTGTCTAACTCTGTAACCTGTGTAACCTTTCCAGACAGTTTTAAAT 744
Qy      48  ProThrTyrAsnAsnHisLeuTyrLysGlnIleSerSerAlaSerThrGlyAlaSerAsn 67
Db      745  CCATATATGCCAGACCACTTATAGGTGTGTTCTCCCGAGCGAGTAGCTGCCCAAT 804
Qy      68  Asp-----AsnHisTyrPheGlyTyrSerThrPro 77
Db      805  GCCAGTGGAAAGGAGGCAAGGTTTGCCACCATCAGTCCCATATATGGGATACTCAACCCCA 864
Qy      78  TrpGlyTyrPheAspPheAsnArgPheHisCysHisPheSerProArgAspTrpGlnArg 97
Db      865  TGGAGATATTAGATTTTAAATGCTTTAAATTTATTTTTCACCTTTAGAGTTTCAGCAC 924

```

```

Qy      98  LeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAsnIle 117
Db      925  TTAATTGAAATTTATGGAAGTATAGTCTCTGATGCTTTAACTGTAACTGTAACTATCAGAAAT 984
Qy      118  GlnValLysGluValThr-----ThrAsnAspGlyValThrThrIleAlaAsnAsnLeu 135
Db      985  GCTGTTAGGATGTTTACAGACAAACTGGAGGGGGGTA---CAGGTTACTGACAGCACT 1041
Qy      136  ThrSerThrValGlnValPheSerAspSerGluTyrGlnLeuProTyrValLeuGlySer 155
Db      1042  ACAGGGCCCTATGCTATGATTAGTACCATGAATACAAAGTACCCATATGTGTAGGGCAA 1101
Qy      156  AlaHisGlnGlyCysLeuProPheProAlaAspValPheMetIleProGlnTyrGly 175
Db      1102  GGTCAAGATACCTTAGCCCCAGAACTCTATTGGGTATACTTTCCCCCTCAATATGCT 1161
Qy      176  TyrLeuThrLeuAsnAsn---GlySerGlnAlaValGly----- 187
Db      1162  TACTTAACAGTAGGAGATGTTAAACACACAGGAATTTCTGGAGACAGCAAAATTAGCA 1221
Qy      188  -----ArgSerSerPheTyrCysLeuGluTyrPheProSerGlnMetLeuArgThrGly 205
Db      1222  AGTGAAGAATCAGCACTTTTATGTTTGAACACACAGTCTCTTTTCAGCTTTTAGGTACAGA 1281
Qy      206  AsnAsnPheThrPheSerTyrThrPheGluGluValProPheHisSerSerTyrAlaHis 225
Db      1282  GGTACAGCACTATGCTCTTATAAGTTTCTCCAGTGCCTCCAGCAAAATTTAGAGGGTGC 1341
Qy      226  SerGlnSerLeuAspArgLeuMetAsnProLeuIleAspGlnTyrLeuTyrTyrLeuAsn 245
Db      1342  AGTCAACACTTTTATGAAATGTACAATCCCTTA-----TAGGGATCCCGCTTAGGG 1392
Qy      246  ArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeuPheSerArgGlySer 265
Db      1393  GTTCTCTGACACATTAGGAGGTGACCCAAATTTAGATCTTTA-----ACA 1437
Qy      266  ProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyProCysTyrArgGlnGln 285
Db      1438  CATGAAGACCATGCAATTCAGCCCCCAAACTTCATGCCAGGCCCACTAGTAACTCAGTG 1497
Qy      286  ArgValSerLysThrLysThrAspAsnAsnAsnSerAsnPheThrTrpThrGlyAlaSer 305
Db      1498  TCTACAAAGGAGGAGACAGCTCTAATACTGAGAGTGGAAAGCCCTTAACAGGCTTAGC 1557
Qy      306  LysTyrAsnLeuAsnGlyArgGluSerIleIleAsnProGly---ThrAlaMetAlaSer 324
Db      1558  ACAGGTACCTCTCAAAACACTAGAATATCTTTACGCCCTGGGCCAGTGTCTCAGCCATAC 1617
Qy      325  HisLysAspAspGluAspLysPhePheProMetSerGlyValMetIlePheGlyLysGlu 344
Db      1618  CACCCTGGACACAGATAAATATGTCACAGGAATAAATGCCATTTCTCATGTCAGACCC 1677
Qy      345  SerAlaGlyAlaSerAsnThrAlaLeuAspAsnVal-----Met 357
Db      1678  ACTTATGGT-----AACGCTGAAGACAAAGAGTATCAGCAAGGAGTGGGTAGATTT 1728
Qy      358  IleThrAspGluGluLeuIleLysAlaThrAsnProValAlaThrGluArgPheGlyThr 377
Db      1729  CCAATGAAAGAGACAGCTAAACAGATTACAGGGTTTAAACATGCACACCTAC----- 1782
Qy      378  ValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHisAlaMet 397
Db      1783  -----TTTCCCAATAAAGGNAACCCAGCAATATACAGATCAAAATTTGAG---CGC 1827
Qy      398  GlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyProIleTrp 417
Db      1828  CCCCTAATGGTGGTCTCTGTATGGAACAGAGAGGCCCTTCACTATGAAAGCCAGCTGTG 1887
Qy      418  AlaLysIleProHisThrAspGlyHisPheHisProSer---ProLeuMetGlyGlyPhe 436
Db      1888  AGTAAATTTCAAAATTTAGTACAGTTTTAAACCTCAGTTTGCAGCCTTAGGAGGATGG 1947

```


| | | | |
|----|------|--|------|
| Qy | 437 | GlyLeuLysAsnProProGlnIleLeuIleLysAsnThrProValProAlaAsnPro | 456 |
| | | | |
| | | | |
| Db | 1948 | GGTTTGATCAGCACCTCTCAATATATTTTAAAA-----ATATTACCACAAAGTGGG | 2001 |
| | | | |
| | | | |
| Qy | 457 | ProAlaGluPheSerAlaThrLysPheAlaSerPheIleThrGlnTyrSerThrGlyGln | 476 |
| | | | |
| | | | |
| Db | 2002 | CCAATTGGAGGTATTAATCAATGGGNAATTACTACTTAGTTCAGTATCCGTGGGNAATT | 2061 |
| | | | |
| | | | |
| Qy | 477 | ValSerValGluLeuGluTrpGluLeu---GlnLysGluAsnSerLysArgTrpAsnPro | 495 |
| | | | |
| | | | |
| Db | 2062 | ATGACAGTAACATGACATTTAAATTGGGGCCCGTAAAGCTACGGGACGGTGGAAATCCT | 2121 |
| | | | |
| Qy | 496 | Glu | 496 |
| | | | |
| Db | 2122 | CAA | 2124 |

Search completed: February 15, 2004, 01:39:33
Job time : 192.857 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 14, 2004, 15:26:50 ; Search time 2963.43 Seconds
(without alignments)
4379.586 Million cell updates/sec

Title: US-09-807-802A-17
Perfect score: 2906
Sequence: 1 MASGGGAPWADNNEGADGVG.....NNGLYTEPRPTGTRYLTRPL 534

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh
-O=/cpn2_1/USPTO_spool/US09807802/runat_11022004_175608_15941/app query.fasta_1.2389
-DB=EST-QPMT-fastap -SUFFIX=rest -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09807802 @CNC 1 1 7257 @runat_11022004_175608_15941 -NCPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_esthum:*
4: em_esthum:*
5: em_esthum:*
6: em_estpl:*
7: em_estro:*
8: em_hc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_ges_hum:*
18: em_ges_inv:*
19: em_ges_pln:*
20: em_ges_vrt:*
21: em_ges_fun:*
22: em_ges_mam:*
23: em_ges_mus:*
24: em_ges_pro:*
25: em_ges_rod:*
26: em_ges_pbg:*
27: em_ges_vrl:*
28: gb_ges1:*

29: gb_ges2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 134 | 4.6 | 856 | 28 | BH164736 |
| 2 | 125 | 4.3 | 754 | 10 | EG207226 |
| 3 | 121.5 | 4.2 | 925 | 28 | BH152154 |
| 4 | 112 | 3.9 | 753 | 28 | BH115587 |
| 5 | 110.5 | 3.8 | 768 | 14 | CD558053 |
| 6 | 110.5 | 3.8 | 815 | 10 | EG746848 |
| 7 | 110.5 | 3.8 | 5809 | 11 | AK076994 |
| 8 | 109 | 3.8 | 687 | 13 | BH723756 |
| 9 | 106.5 | 3.7 | 754 | 10 | BE908188 |
| 10 | 106.5 | 3.7 | 786 | 10 | BE733520 |
| 11 | 106.5 | 3.7 | 830 | 13 | BQ888825 |
| 12 | 106.5 | 3.7 | 986 | 13 | BQ682489 |
| 13 | 105 | 3.6 | 606 | 13 | BH800802 |
| 14 | 103.5 | 3.6 | 775 | 12 | B1253826 |
| 15 | 103.5 | 3.6 | 865 | 10 | BF340815 |
| 16 | 103 | 3.5 | 906 | 13 | BH848986 |
| 17 | 102.5 | 3.5 | 626 | 13 | BQ416589 |
| 18 | 102.5 | 3.5 | 2742 | 11 | AK004837 |
| 19 | 102 | 3.5 | 829 | 13 | BH365695 |
| 20 | 102 | 3.5 | 2637 | 11 | BC037597 |
| 21 | 102 | 3.5 | 2688 | 11 | BC019128 |
| 22 | 101.5 | 3.5 | 645 | 12 | BM771938 |
| 23 | 101 | 3.5 | 838 | 28 | AZ686611 |
| 24 | 101 | 3.5 | 3057 | 11 | AK044947 |
| 25 | 100 | 3.4 | 688 | 12 | BM156442 |
| 26 | 100 | 3.4 | 959 | 14 | CD458291 |
| 27 | 100 | 3.4 | 962 | 13 | BQ684616 |
| 28 | 100 | 3.4 | 1201 | 13 | BX442505 |
| 29 | 100 | 3.4 | 2036 | 10 | BF128255 |
| 30 | 100 | 3.4 | 4479 | 11 | BC044003 |
| 31 | 99.5 | 3.4 | 672 | 29 | CC198495 |
| 32 | 99.5 | 3.4 | 691 | 29 | BZ315618 |
| 33 | 99.5 | 3.4 | 739 | 14 | CA498545 |
| 34 | 99.5 | 3.4 | 852 | 14 | CB990709 |
| 35 | 99.5 | 3.4 | 894 | 12 | B1759543 |
| 36 | 99.5 | 3.4 | 950 | 13 | BH172290 |
| 37 | 99 | 3.4 | 871 | 10 | BF269841 |
| 38 | 99 | 3.4 | 912 | 13 | BX325591 |
| 39 | 99 | 3.4 | 1039 | 29 | CNS077AT |
| 40 | 99 | 3.4 | 2029 | 11 | AY104007 |
| 41 | 99 | 3.4 | 3329 | 11 | AK033012 |
| 42 | 98.5 | 3.4 | 841 | 13 | BX392614 |
| 43 | 98.5 | 3.4 | 932 | 13 | BH838941 |
| 44 | 98.5 | 3.4 | 1201 | 9 | AL520051 |
| 45 | 98 | 3.4 | 704 | 9 | AL043249 |

ALIGNMENTS

RESULT 1
BH164736
LOCUS
DEFINITION
ENTTS28TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, genomic survey sequence.
ACCESSION
BH164736
VERSION
BH164736.1
KEYWORDS
GSS.
SOURCE
Entamoeba histolytica
ORGANISM
Entamoeba histolytica
REFERENCE
1 (bases 1 to 856)
AUTHORS
Lofthus,B., Wang,Z., Van Aken,S. and Fraser,C.

BH164736 856 bp DNA linear GSS 24-SEP-2001
ENTTS28TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, genomic survey sequence.

BH164736 GI:15738174
GSS.
Entamoeba histolytica
Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
1 (bases 1 to 856)
Lofthus,B., Wang,Z., Van Aken,S. and Fraser,C.

TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library (2001)
COMMENT Unpublished
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Seq primer: M13-Forward
 Class: shotgun
 High quality sequence start: 17
 High quality sequence stop: 629.
FEATURES Location/Qualifiers
 source
 1..856
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: PHOS1; Site 1: Best I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Barrell, Oxford University Press, 1999)."

BASE COUNT 347 a 229 c 89 g 191 t
ORIGIN

Alignment Scores:
 Pred. No.: 0.0039 Length: 856
 Score: 134.00 Matches: 88
 Percent Similarity: 33.42% Conservative: 44
 Best Local Similarity: 22.28% Mismatches: 140
 Query Match: 4.61% Indels: 123
 DB: 28 Gaps: 19

US-09-807-802a-17 (1-534) x BH164736 (1-856)

Qy 95 TtpGlnArgLeuIleAsnAnanTtpGlyPheArgProLysArgLeuAsnPhelLysLeu 114
 Db 9 TGGAGACACTGTTGGTAAAT-----TCATTCCTGCTGGA 47

Qy 115 PheAsnIleGlnValLysGluValThrAsnAspGlyValThr----- 130
 Db 48 TTTAATACTAACCAACGTCGGGAACAATAATCCGTTCAATTCCTACTGGAATTCAGTC 107

Qy 131 ----IleAlaAsnAsnLeuThrSerThrValGlnValPheSerAspSerGlnLeu 149
 Db 108 TCAGGAGCAATAACCTTTTGTCTACTACTCA-----AACAAATACAACT 155

Qy 150 ProTyrValLeuGlySerAlaHisGlnGlyCysLeuProProPheProAlaAspValPhe 169
 Db 156 CTTTCACACAGGACAACTAAC-----CCATTTAAT----- 188

Qy 170 MetIleProGlnTyrGlyTyrLeuThrLeuAsnAsnGlySerGlnAlaValGlyArgSer 189
 Db 189 -----ACTACTAACAAATACTACTACTCA----- 212

Qy 190 SerPheTyrCysLeuGlnTyrPheProSerGlnMetLeuArgThrGlyAsnAsnPhetr 209
 Db 213 -----ACAACTAATCACTTTAAT 230

Qy 210 PheSerTyrThrPheGluGluValProPheHisSerSerTyrAlaHisSerGlnSerLeu 229
 Db 231 ACTAATACTACAGCACCACCAATCAATTTAATACAATACTACTAAT----- 278

Qy 230 AspArgLeuMetAsnProLeuIleAspGlnTyrTyrLeuAsnArgThrGlnAsn 249
 Db 279 -----AATCCATTT-----AGTACTAATAAAC 299

Qy 250 GlnSerGlySerAlaGlnAsnLysAspLeuLeuPheSerArgGlySerProAlaGlyMet 269
 Db 300 ACTACCAATGAAACAACCAAT-----TTATTAAATCAA-----ACTCCTTCTCTAAT 347

Qy 270 SerValGlnProLysAsnTtpLeuProGlyProCysTyrArgGlnGlnArgValSerLys 289
 Db 348 ACTATAACACCCAGGAAT-----AATACAACACTACT 377

Qy 290 ThrLysThrAspAsnAsnAsnSerAsnPhetrTtpThrGlyAlaSerLysTyrAsnLeu 309
 Db 378 ACAACAGGAAGTAATCACTTTTGGTAATTTCACTACCCTAAT----- 419

Qy 310 AsnGlyArgGluSerIleIleAsnProGlyThrAlaMetAlaSerHisLysAspAspGlu 329
 Db 420 AATACTACTCTCCAGTACAACCTTCAACAGGAACAACCACTACACAGGAAGTAATCCATTT 479

Qy 330 AspLysPhePheProMetSerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSer 349
 Db 480 GGTAAATTTCACTCTCCCTCAACTCAGCTCCACACAGGAATAATAACACAGGAACAGCT 539

Qy 350 AsnThrAlaLeuAsnValMetIleThrAspGluGluGluIleLysAlaThrAsnPro 369
 Db 540 ACAACACAGGAATAATAACCACTCACA-----ACAGGAAGTAATCA 584

Qy 370 ValAlaThrGluArgPheGlyThrValAlaValAsnPheGlnSerSerSerThrAspPro 389
 Db 585 -----TTTGGT-----AATTCATCTCCTCAACACCACTCCC 617

Qy 390 AlaThrGlyAspValHisAlaMetGlyAlaLeuProGlyMetValTtpGlnAspArgAsp 409
 Db 618 ACCACAGGAAT-----AATACACAGGAACAACCTACCAACAGGAAT 662

Qy 410 ValTyrLeuGlnGlyProIleTtpAlaLysIleProHisThrAspGlyHisPheHisPro 429
 Db 663 AAATACACAGGA-----ACAGGAAGTATCCATTT-----GGTAATTTCACTACT 707

Qy 430 SerProLeuMetGlyGlyPheGlyLeuLysAsnProProProGlnIleLeuIle-LysAs 449
 Db 708 TCAAACTCACTACCCACACAGGAACAACAACCTACACAGGAATAACACACTACAGGA 767

Qy 449 nThrProValProAlaAsnProProAlaGluPheSerAlaThrLysPheAlaSerPheIle 469
 Db 768 CAACCCCAACCGGGAACACACCCCAACCGGAACACCCCAAC----- 810

Qy 469 eThrGlnTyrSerThrGlyGlnValSerValGluIleGluTtp 483
 Db 811 -----ACAGGAAGGATCCTTTTGGAAATTTTCATAC 840

RESULT 2
 BG207226 754 bp mRNA linear EST 21-APR-2001
 LOCUS R5226594 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
 DEFINITION BG207226
 ACCESSION BG207226.1 GI:13728913
 VERSION 1
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 754)
 Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,
 Cain, S., Leventhal, C., Thornton, M., Ramachandran, R., Whittington, J.,
 Lerner, L., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith,
 E., Veloso, N., Klika, A., Hess, J., Cothren, K., Lo, K., Offenbacher,
 J., Danzig, J. and Ducar, M.

Alignment Scores:

Pred. No.: 0.618 Length: 753
 Score: 112.00 Matches: 25
 Percent Similarity: 59.32% Conservative: 10
 Best Local Similarity: 42.37% Mismatches: 24
 Query Match: 3.85% Indels: 1
 DB: 28 Gaps: 0

US-09-807-802A-17 (1-534) x BH115587 (1-753)

Qy 1 MetAlaSerGlyGlyGlyAlaProMetAlaAspAenAenGluGlyAlaAspGlyValGly 20
 Db 521 ATCTCTGGAGTGGAGGCTTACTAATGGGCAACAACACGAGC-ACGTGATGGAGTGAGC 579
 Qy 21 AsnAlaSerGlyAsnTrpHisCysAspSerTrpTrpLeuGlyAspArgValIleThrThr 40
 Db 580 AATGCCATGGATGGATGGCATGGCCATCCCAAGTGTATGGTAGCTGAGTCATTACCTAA 639
 Qy 41 SerThrArgThrTrpAlaLeuProThrTyrAenAenHisLeuTyrLysGlnIleSer 59
 Db 640 TCTCCCAAGACCTGGGCTCTGCCCAACTCCCACTAACATTTCCACACTATATGAAC 696

RESULT 5

CD558053 768 bp mRNA linear EST 11-JUN-2003
 LOCUS AGENCOURT 14412765 NIH_MGC_180 Homo sapiens cDNA clone
 DEFINITION IMAGE:30386880 5', mRNA sequence.

ACCESSION CD558053.1 GI:31584121

VERSION

EST. Homo sapiens (human)

KEYWORDS

Homo sapiens

SOURCE

Mammalia; Eutheria; Primates; Craniata; Vertebrata; Euteleostomi;

Eukaryota; Eutheria; Primates; Craniata; Vertebrata; Euteleostomi;

National Institutes of Health, Mammalian Gene Collection (MGC)

REFERENCE

1 (bases 1 to 768)

NIH-MGC http://mgi.nci.nih.gov/.

AUTHORS

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished

COMMENT

Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics / NIH
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Dr. Michael Brownstein
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: NDAM456 row: e column: 01
 High quality sequence stop: 444.

FEATURES

source

1..768
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30386880"
 /lab_host="PH108-Ton A (T1 and T5 phage resistances)"
 /clone_lib="NIH-MGC_180"
 /note="Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI;
 Site 2: EcoRV (destroyed); Library is oligo-dT primed and
 directionally cloned (EcoRV site is destroyed upon cloning
). Average insert size 1.68 kb. Library was constructed by
 (Invitrogen). Note: this is a NIH-MGC Library."
 BASE COUNT 178 a 216 c 216 g 158 t

ORIGIN

Alignment Scores:
 Pred. No.: 0.912 Length: 768
 Score: 110.50 Matches: 62
 Percent Similarity: 36.36% Conservative: 22
 Best Local Similarity: 26.84% Mismatches: 69

Query Match: 3.80% Indels: 79
 DB: 14 Gaps: 14
 US-09-807-802A-17 (1-534) x CD558053 (1-768)
 Qy 271 ValGlnProLysAenTrpLeuProGlyProCysTyrArgGlnGlnArgValSerLysThr 290
 Db 182 ATTCAAGCCCGATGGTCAATGCCCAAGT-:::|||||-----GATAAAACC 217
 Qy 291 -----LysThrAspAenAenSerAenPheThrTrpTrpGlyAlaSerLysThr 307
 Db 218 ATTGGTCGGGGACGACTCTCTCAACAGCTTCTTCACTGAGACTGAGCTGGCAAGC 277
 Qy 308 AsnLeuAenGlyArgGluSerIleAenProGlyThrAlaMet-----AlaSer 324
 Db 278 GT-----GCCACAGCAGCTGTTTGTGGACCTGGAGCCCACTGGTGGTCAAGTGC 330
 Qy 325 HisLysAspAspGlu-AspLysPhePhePrometSerGlyValMetIlePheGlyLysG 344
 Db 331 CACAGGACCTACAGCAGCTCTTCCACCCACAGCAG-----CTGATCACCGGAGGA 384
 Qy 344 uSerAlaGlyAlaSerAenThrAlaLeuAspAenValMetIleThrAspGluGlu 364
 Db 385 AGATGCA---GCCAGTAATTACGCCAGGGCCATTACACCATCGCAAGGAGATTGTGA 441
 Qy 364 eLysAlaThrAenProValAlaThrGluArgPheGlyThrValAlaValAenPheGlnSe 384
 Db 442 CCTAGTCTCGGACCG-----:::|||||:::|||||:::|||||:::||||| 457
 Qy 384 rSerSerThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeuProGlyMetVa 404
 Db 458 -----GATCCGCAAACTGGGCGGATCTGTGCACAGGA---CTGCCAGGCTTTC 504
 Qy 404 1-----TTPGlnAsp-----ArgAspValTY 411
 Db 505 CTGATCTTTCACAGCCTTTTGGGGGGCGGCAACTGGCTTCTGGGGTTCGATCTCTG 564
 Qy 411 rLeuGlnGly-----:::|||||:::|||||:::|||||:::||||| 414
 Db 565 GCTCATGGGAAACGGGCTCTTTCAGTGGGAATTACCAGCAAAAGAAAGTTCGAAACCTAA 624
 Qy 415 -ProletTrpAlaLysIleProHisThr-----AspGlyHisPheHisProSerProle 432
 Db 625 GAGTTTTTGGCCATTTTACCCCAAGCCCCCAAGGGGTTCTTCCACCAACCCCTT 684
 Qy 432 uMetGlyGlyPheGlyLeuLysAenProProGlnIleLeuIle-----LysAs 449
 Db 585 G-----GGTTGGAAACCCCTTAAACAACTTCCCATTCCTCCCTAAACAAAC 732
 Qy 449 nThrProValProAlaAenProProAlaGlu 459
 Db 733 CACCCCAAAAGGAAACCCCTTGGAA 763

RESULT 6

BG746848/c

LOCUS

602704187F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4857428 5',

DEFINITION

mRNA sequence.

ACCESSION

BG746848

VERSION

BG746848.1 GI:14057501

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 815)

NIH-MGC http://mgi.nci.nih.gov/.

AUTHORS

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: NIH Intramural Sequencing Center
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LLCMI709 row: n column: 21
 High quality sequence stop: 809.

FEATURES

Location/Qualifiers

1..815
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4857428"
 /issue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 15"
 /note="Organ: colon; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGACGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)"
 BASE COUNT 176 a 246 c 243 g 149 t
 ORIGIN

Alignment Scores:

Pred. No.: 0.998 Length: 815
 Score: 110.50 Matches: 69
 Percent Similarity: 36.74% Conservative: 46
 Best Local Similarity: 22.04% Mismatches: 124
 Query Match: 3.80% Indels: 74
 DB: 10 Gaps: 11

US-09-807-802A-17 (1-534) x BG746848 (1-815)

QY 197 PheProSerGlnMetLeuArgThrGlyAsnAsnPheThrPheSerTyrrhrPheGluGlu 216
 Db 802 TTCGGCAACGAGGAGAGAGAACTAGTCTGAGTGTGTTTGTAGTGTGTTTCTAGAA 743
 QY 217 ValProPheHisSerSerTyrrAlaHisSerGlnSerLeu-----AspArgLeuMet 233
 Db 742 ATAGGACATATTGGTGTGTTTAAACAATTAATAATCTTCTGGCAGCCCTTTGCTC 683
 QY 234 AsnProLeuLeuAspGlnTyrrLeuTyrrLeuAsnArgThrGln----- 248
 Db 682 AGGCTGTGTGCTCT-----GCTCAGCCACTCTGTCCACCACTGACCCAGGGGCTGC 629
 QY 249 -----AsnGlnSerGlySerAlaGlnAsnLysAspLeuLeuPheSerArgGlySerPro 266
 Db 628 CCGGACATAGACGAGGGGGTTCGGGCCCACTCCCTGTCATGCCGCCACCCACCC 569
 QY 267 AlaGlyMetSerValGlnProLysAsnTrpLeuProGlyProCysTyrrArgGlnArg 286
 Db 568 GGGCTGTGCCACTACAGCCGGCTCACTGGCTCCAGGGACCACTCCAGCCATACAG 509
 QY 287 ValSerLysThrLysThrAspAsnAsnSerAsnPheThrTrpThrGlyAlaSerLys 306
 Db 508 CTTTCATCTGTGCCAGGTCCCTCTGTGTCAAAGCTCTTGTGTATGGCCAGGG----- 458
 QY 307 TyrAsnLeuAsnGlyArgGluSerIleAsnProGlyThrAlaMetAlaSerHis-Ly 326
 Db 457 -----CCTGCGACAGCTTTCGAGGGGCCACTGC 431
 QY 326 sAspAspGluAspLysPhePhePrometSerGlyValMetIlePheGlyLysGluSerAl 346
 Db 430 AGAG-----CTCGCGTCTCTTCGAGACACGGC 404
 QY 346 aGlyAlaSerAsnThrAlaLeuAspAsnValMetIleThrAspGluGluLeuLysAl 366
 Db 403 TGGCGGTGCC-----TTTCCGAGGGGTCTGCGAGCT 371

QY 366 aThrAsnProValAlaThrGluArgPheGlyThrValAlaValaAsnPheGlnSerSerSe 386
 Db 370 GACGCATGGTCTCAGACC-----TTCACAGGGGTCCAGTCTCGCAGCAGACGCAAG 317
 QY 386 rThrAspProAlaThrGlyAspValHisAlaMetGlyAlaLeuProGlyMetValTrpGl 406
 Db 316 CCCCGTCCCTCGGTGGAGATGTTTCATGCA-----GAAGTACATGC 275
 QY 406 nAspArgAspValTyrrLeuGlnGlyProIleTrpAlaLysIleProHisThrAspGlyHi 426
 Db 274 CGTAAATGATGTACACGTACACGGTCCCGGCTTCATGAACATGCCTCGGTTCGGGGGG 215
 QY 426 sPheHisProSerProLeuMetGlyGlyPheGlyLeuLysAsnProProGlnIleLe 446
 Db 214 TCTCGCGGCCACCTTGTAGTGGGGGCTTCATCTCTGG-----CCCCAGGTATG 164
 QY 446 uIleLysAsnThrProValProAlaAsnProProAlaGluPheSerAlaThrLysPheAl 466
 Db 163 CCTC-----GGTCTCCACGATCGGCTCGGAGTTCTG 131
 QY 466 aSerPhe-----IleThrGlnTyrrSerThrThrGlyGlnValSerVa 479
 Db 130 TGCCATTAGGAAGTCGCGGACTAGGACCTGTCCAGAAATGCCCGGCGCAGGGGACTG 71
 QY 479 lGluIleGluTrpGluLeuGlnLysGluAsnSerLys 491
 Db 70 CCGGCTGTGCGAAGAACTCCAAACCCAGTCGGGTAAAG 34

RESULT 7
 AK076994
 LOCUS
 DEFINITION

AK076994 5809 bp mRNA linear HTC 07-DEC-2002
 Mus musculus adult male testis cDNA, RIKEN full-length enriched
 library, clone:4931406P16 product:hypothetical protein, full insert
 sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

AK076994
 GI:26345807
 HTC; CAP trapper.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

1
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 99279253
 10349636

2
 Carninci, P., Shibata, Y., Hayata, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 20499374
 11042159

3
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujikawa, S., Inoue, K., Togawa, Y., Iizawa, M., Ohara, E., Wataniki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861

4
 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
 Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
 Aizawa, K., Iizawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
 Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,

Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Mateu, Y., Nikaide, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilmberg, D., Wyndham-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y.

TITLE
Functional annotation of a full-length mouse cDNA collection
JOURNAL
Nature 409 (6821), 685-690 (2001)
MEDLINE
21085660
PUBMED
11217851
REFERENCE
5

AUTHORS
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL
Nature 420, 563-573 (2002)
REFERENCE
6 (bases 1 to 5809)

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurthara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saigo, H., Saito, R., Sakai, K., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, P., Tanaka, T., Tejima, F., Toyota, T., Yamamura, T., Yamanaka, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

TITLE
Direct Submission
JOURNAL
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-resesc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/

FEATURES
source

```

1..5809
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:4931406P16"
/db_xref="MGI:1910824"
/db_xref="taxon:10090"
/clone="4931406P16"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
415..3594
/note="unnamed protein product; hypothetical protein (evidence: rscDS, ProCrest, decoder, NCBI CDS Predictor, Longest-ORF) putative"
/codon_start=1

```

CDS

```

/protein_id="BAC36555.1"
/db_xref="GI:26345808"
/db_xref="MGI:1924311"
/translation="MYCSQAQNDKMDYKRRFLGSGSKOKVOHQOYPMPELGRALSAP
LASTAPAPGSLTAAGSCHAMPHSTPIADIQQGISKYLDALINVPFCRASTPLTDLPS
TVFNRSYKAAQMLQDVQEHVMEARSLTSARPEIAOMLMEAGAAATFDQKEFS
LQDEIVLGRCLTVVQVHFQPLTHALQKVQVFAEVVVPKONSLSLGMGHTPE
LEEAVRWRGAAEATSLRERGCDCGLAGIEVQOLFCSQSAAIPEHQJKEINIKDIA
LQAYKILGSHLCHEYAKAGFLNPKAIEASLQCCSEAEAAQOTGRRTQPPMQCE
LPTVPVQIIGSHFLKGVSNESAADNLKLTHTMLQLIKEACYNIGITPRDDLPTVEVL
NOVCPSWRGACKTAVOLLFQOAGLVVDTAQIENKAYAPQISLEGSRIYVQVPSIW
CLFGDPATMLLSRLOPEKLTGLVDLITAVLDLNRWRTGREALPCIQIQLQKDC
DKENQADPLSGNKSGLQKTFSLKTSRFTKVCSTNSGTHYSKNLFTAGCSERA
KMMNNITDRLQISLITGNFRTDPSQSSNSPMVNGFLRRERENFLRGGDQKES
KGMNLTQDEMVEIDFLSGFMQSGHQSPLVTRNSTATAMVTEQKTGTWQPOQS
LPLAPPLRPOAGAHTPLAAQOGLAPOQSPQOQVVOYHLLQPTIGSOOTPOPR
APAKWHGSSQHPSPQMAAGLPLGOMGGLSDLSGLSVSYNDNMMSVLGOK
POGPRNNTWNRDQSDGVFGMLGEILFPDPAVGDSPFARYAVGVSQMQKRAQHG
RRPGRNPNPMMEDAHRTWPLPEFFTEGSLHSGWGAQGSASSDETSSANGSL
FSMFSGPLVAQKRRKSGEQESTLPSPLLTVDVNDKNTKTPPKAPMQH
PSPMPSTLPSPAAPLYAVAGPSQWMDTVQMLQSPVMAAANDCNTTSYVVTPTPPQ
PPPAHKAAPKGEKAFPGKAERRPAYLPQY"
5794..5799
/note="putative"
5809
/note="putative"
polyA_signal
polyA_site
BASE COUNT 1531 a 1471 c 1362 g 1445 t
ORIGIN
Alignment Scores:
Pred. No.: 19.6 Length: 5809
Score: 110.50 Matches: 108
Percent Similarity: 33.60% Conservative: 63
Best Local Similarity: 21.22% Mismatches: 187
Query Match: 3.80% Indels: 155
DB: 11 Gaps: 22
US-09-807-802A-17 (1-534) x AK076994 (1-5809)
QY 6 GlyAlaProMetAlaAapAsnAsnGluGlyAlaAapGlyValGlyAsnAlaSerGlyAen 25
Db 2344 GCCTTTAACTGGGCCAGTCCCATCAGGGGTCCCATGAGCGGTAACTCTACT 2403
QY 26 TrpHisCysAapSerThrTrpLeuGlyAapArg-----Valle 38
Db 2404 GC-CACCTGCCATGGTGAAGAACAGAGACAGGAACCAACACAGCAGCAGCTACT 2462
QY 39 ThrThrSerThrArgThrTrpAlaLeuProThrThrAsnAsnHisLeuTyLysGlnle 58
Db 2463 GCCATTGGCTCTCCACTACGCCGCCACCCAGCGCTGGAGACACACAGCTCTGG-CAGCCC 2521
QY 59 Ser-----SerAlaSerThrGlyAlaSerAsnAsnHisTyPheGly 73
Db 2522 AGCAGGTCTGGCTCCTCAGCAGCAGCTCTCAAGCAGCAGCAGCCCCCAAGTCCAGTACT 2581
QY 74 TyrSerThrProTrpGlyTyPheAspPheAsnArgPheHis----- 87
Db 2582 ACCAGCACCTGCTGCAACCCATTGGATCAGCAGAGACTCCACCTCAGCTCGGGCGCTG 2641
QY 88 -----CysHisPheSerProArgAapTrpGlnArg--- 97
Db 2642 CGAATGGGTACATGGCTCATCTCAGCAGCACCCTCAGCCCA-----TGCAGCGGGTCT 2695
QY 98 ---LeuIleAsnAsnAsnTrpGlyPheArgProLysArgLeuAsnPheLysLeuPheAen 116
Db 2696 TGTCTCTCTCTT-GTTCAGTGG----- 2715
QY 117 IleGlnValLysGluValThrThrAsnAspGlyValThrIleAlaAsnAsnLeuThr 136
Db 2716 -----CCTGGCATATCTGATCTCAGCTCTGACTGTGAC 2748
QY 137 SerThrValGlnValPheSerAspSerGluTyGlnLeuProTyValLeuGlySerAla 156

```


Qy 216 -----GluValProPheHisSerSerTyrAlaHis 225
 Db 414 TTACAGAAATTGAATATTGAATCGTGCACGTATACCATATCATCGT----- 461
 Qy 226 SerGlnSerLeuAspArgLeuMetAsnProLeuLeuAspGlnTyrLeu---TyrTyrLeu 244
 Db 462 -----CAACAATATGTAACAATATATGTT 485
 Qy 245 AsnArgThrGlnAsnGlnSerGlySerAlaGlnAsnLysAspLeuLeuPheSerArgGly 264
 Db 486 AATAATAATAATAATGATAAT-----ATTCTGAAGAATATACGTTTTCACAT--- 533
 Qy 265 SerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyProCysTyrArgGln 284
 Db 534 CCTCAAGCAACAATTTCTGTAAAT-----TATCAACAC 566
 Qy 285 GlnArgValSerLysThrLysThrAspAsnAsnAsnSer 297
 Db 567 CAACAACAACATAATAATAATAATGATAATAATAATAAT 605

RESULT 9
 BE908188/c
 LOCUS 601500441F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3902123 5',
 DEFINITION mRNA sequence.
 BE908188
 VERSION BE908188.1 GI:10402511
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 754)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM9704 row: b column: 12
 High quality sequence stop: 678.
 Location/Qualifiers
 1..754
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3902123"
 /tissue_type="epithelioid carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_70"
 /note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.1 kb. Library constructed by Life
 Technologies."

FEATURES

source
 153 a 231 c 239 g 131 t
 BASE COUNT
 ORIGIN
 Alignment Scores:
 Pred. No.: 2.31 Length: 754
 Score: 106.50 Matches: 56
 Percent Similarity: 36.65% Conservative: 36
 Best Local Similarity: 22.31% Mismatches: 97
 Query Match: 3.66% Indels: 62
 DB: 10 Gaps: 8

US-09-807-802A-17 (1-534) x BE908188 (1-754)

Qy 249 AsnGlnSerGlySerAlaGlnAsnLysAspLeuLeuPheSerArgGlySerProAlaGly 268
 Db 594 CATAGAAGCGGAGGGGTTTCCGGGCCCATCTCCCTGTCATGGCGGACGCCACCGGGCTG 535
 Qy 269 MetSerValGlnProLysAsnTrpLeuProGlyProCysTyrArgGlnArgValSer 288
 Db 534 CTGCCCTACAGCGGCTCACTGGCTCCAGGGGACCACCTCCAGGCATACAGCTTCAT 475
 Qy 289 LysThrLysThrAspAsnAsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsn 308
 Db 474 CCTGTGCAGGTCCTCTGGTCAAGGCTCTGTTGATGGCCAGG----- 430
 Qy 309 LeuAsnGlyArgGluSerIleAsnProGlyThrAlaMetAlaSerHis-LysAspAs 328
 Db 429 -----CCTGGCACACGCTTGGAGGGGCCCATCGCAGAG-- 399
 Qy 328 pGluAspLysPhePheProMetSerGlyValMetIlePheGlyLysGluSerAlaGlyAl 348
 Db 398 -----CTCGGGTCTTGTGAGGACACGGCTGGCGG 370
 Qy 348 aSerAsnThrAlaLeuAspAsnValMetIleThrAspGluGluGluIleLysAlaThrAs 368
 Db 369 TGCC-----TTTCCGGAGGGTGTCTCGGAAGCTGACGCCA 337
 Qy 368 nProValAlaThrGluArgPheGlyThrValAlaValAsnPheGlnSerSerSerThrAs 388
 Db 336 TGGTCTCCAGACC-----TTCCAGGGGCTCCAGTGTCTCGCAGCAGCAGCAAGCCCGCT 283
 Qy 388 pProAlaThrGlyAspValHisAlaMetGlyAlaLeuProGlyMetValTrpGlnAspAr 408
 Db 282 CCCCTGGCTGGAGATGTTTCATGCA-----GAAGTACATGCGCGTAA 241
 Qy 408 gAspValTyrLeuGlnGlyProIleTrpAlaLysIleProHisThrAspGlyHisPheH 428
 Db 240 TGATGTACACGTACAGGGTCCCGGCTTCATGAACATGCTCGGTTCGGGGGGTCTGCCC 181
 Qy 428 sProSerProLeuMetGlyPheGlyLeuLysAsnProProGlnIleLeuIleLys 448
 Db 180 GGCCACCCCTTGAGTGGGGGCTTCATCCTCTGG-----CCCCAGGTATGCCCTC-- 132
 Qy 448 aAsnThrProValProAlaAsnProProAlaGluPheSerAlaThrLysPheAlaSerPh 468
 Db 131 -----GGTCTCCAGCATGCGGCTCGGAGTCTCTGGCAT 97
 Qy 468 e-----lleThrGlnTyrSerThrGlyGlnValSerValGluI 481
 Db 96 TAGGAAGTCGCGGACTAGGACCTGTCCAGAAATCCCGGCCGAGGGGACTGCCGCT 37
 Qy 481 eGluTrpGluLeuGlnLysGluAsnSerLys 491
 Db 36 GGTCAAGAAGTCCCAACCCAGTCGGGTAAAG 6

RESULT 10

LOCUS BE733520/c
 DEFINITION 60156055F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3840973 5',
 mRNA sequence.
 BE733520
 VERSION BE733520.1 GI:10147512
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 786)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LCM531 row: n column: 14
 High quality sequence start: 11
 High quality sequence stop: 772.

FEATURES

source

```

1. 786
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3840973"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC_21"
/notes="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT

ORIGIN

Alignment Scores:

```

Pred. No.:      2.46      Length:      786
Score:          106.50    Matches:      56
Percent Similarity: 36.65%  Conservative: 36
Best Local Similarity: 22.31% Mismatches: 97
Query Match:      3.66%  Indels:      62
DB:              10      Gaps:      8
```

US-09-807-802A-17 (1-534) x BE733520 (1-786)

```

Qy 249 AaGlnSerGlySerAlaGlnAenLysAepLeuLeuPheSerArgLysSexProAlaGly 268
Db 665 CATAGAGCGGAGGGGTTTCGGGCCCCCTCCCTGCATGGCGCGACGCCACCGGGCTG 606
Qy 269 MetSerValGlnProLysAenTrpLeuProGlyProCysTyArgGlnGlnArgValSer 288
Db 605 CTGCCACTACAGCCGGCTCACTGGGCTCCAGGGGACACGCTCCAGCCATACAGCTTCAT 546
Qy 289 LysThrLysThrAspAenAenAenSerAenPheThrTrpThrGlyAlaSerLysTyraen 308
Db 545 CTGTGCCAGGTCCTCTCGTCAAGCTCTTGTGATGCCAGGG----- 501
Qy 309 LeuAaGlyArgGluSerIleAenProGlyThrAlaMetAlaSerHis-LysAspAs 328
Db 500 -----CTGGCACAGCTTGAGGGGCCACTGCAGAG-- 470
Qy 328 pGluAspLysPhePheProMetSerGlyValMetIlePheGlyLysGluSerAlaGlyAl 348
Db 469 -----CTCGCGGTCCTTGAGGACACGCGCTGGCGG 441
Qy 348 aserAenThrAlaLeuAaAspAenValMetIleThrAepGluGluIleLysAlaThrAs 368
Db 440 TGCC-----TTTCGGAGGGGTGTGCGAAGCTGACGCA 408
Qy 368 nProValAlaThrGluArgPheGlyThrValAlaValAaenPheGlnSerSerThrAs 388
Db 407 TGGTCTCCAGACC-----TTCCAGGGGCTCCAGTCTCGCAGCAGACGACGACCCCGT 354
Qy 388 pProAlaThrGlyAspValHisAlaMetGlyAlaLeuProGlyMetValTrpGlnAspAr 408
Db 353 CCCCTCGGTGGAGATGTTTCATGCA-----GAAGTACATACCGCGTAA 312
Qy 408 gApValTyLeuGlnGlyProIleTrpAlaLysIleProHisThrAspGlyHisPheHi 428
Db 311 TGATGTACACGTACAGGGTCCCGGCTTCATGAACATGCTCGTGGTGGGGGGCTGCC 252
Qy 428 sProSerProLeuMetGlyGlyPheGlyLeuLysAaenProProGlnIleLeuIleLys 448
```

```

Db 251 GGCCACCCCTTGAGTGGCGGCTTCATCCTCTGG-----CCCCAGGTATGCTC-- 203
Qy 448 sAenThrProValProAlaAaenProProAlaGluPheSerAlaThrLysPheAlaSerPh 468
Db 202 -----GGTCTCCACGATGCGCGCTCGGAGTTCTGTGCCAT 168
Qy 468 e-----lIleThrGlnTySerThrGlyGlnValSerValGluI 481
Db 167 TAGGAGTCCCGGAGCTAGGACCTGTCGCGAATGCCCGGCCAGGGGACTGCCGGCT 108
Qy 481 eGluTrpGluLeuGlnLysGluAenSerLys 491
Db 107 GGTGCAAGAACTCCACACCCAGTCAGTGGGTAAG 77
```

RESULT 11

BQ888825/c

LOCUS

DEFINITION BQ888825 830 bp mRNA. linear EST 16-AUG-2002
 AGENCOURT 8049690 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6083838
 5', mRNA sequence.

ACCESSION

BQ888825

VERSION

BQ888825.1 GI:22280839

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 830)

NIH-MGC http://mgi.nci.nih.gov/.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbe@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LCM5314 row: k column: 07

High quality sequence stop: 642.

Location/Qualifiers

1..830

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/tissue_type="ductal carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_110"

/note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;

Site_2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Library constructed by

Ling Hong in the laboratory of Gerald M. Rubin (University

of California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH_MGC Library."

BASE COUNT 168 a 248 c 248 g 145 t

ORIGIN

Alignment Scores:

```

Pred. No.:      2.67      Length:      830
Score:          106.50    Matches:      56
Percent Similarity: 36.65%  Conservative: 36
Best Local Similarity: 22.31% Mismatches: 97
Query Match:      3.66%  Indels:      62
DB:              13      Gaps:      8
```

US-09-807-802A-17 (1-534) x BQ888825 (1-830)

Qy 249 AaGlnSerGlySerAlaGlnAenLysAepLeuPheSerArgLysSexProAlaGly 268

```

676 CATAGACGGAGGGGTTTCGGGGCCCACTCCCTGCATGGCCGACGCCACCCGGGCTG 617
Qy 269 MetSerValGlnProLysAsnTrpLeuProGlyProCysTyrArgGlnGlnArgValSer 288
Db 616 CTGCACATACACCGCGCTCCAGTGGCTCCAGGGACCACTGCCTCCAGCCATACAGCTTCAT 557
Qy 289 LysThrLysThrAspAsnAsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsn 308
Db 556 CCTGTGCCAGGTCCTCTGGTCAAGCTCTGTGTGATGTCAGG----- 512
Qy 309 LeuAsnGlyArgGluSerIleLeuAsnProGlyThrAlaMetAlaSerHis-LysAspAs 328
Db 511 -----CCTGTCACAGCTTCAGGGGCACTGCAGAG-- 481
Qy 328 pGluAspLysPhePheProMetSerGlyValMetIlePheGlyLysGluSerAlaGlyAl 348
Db 480 -----CTCGGGTCTCTTGGAGCACGGCTGGCGG 452
Qy 348 aSerAsnThrAlaLeuAspAsnValMetIleThrAspGluGluLeuLysAlaThrAs 368
Db 451 TGCC-----TTTCCGGAGGGTGTCTGCGAAGCTGACGCA 419
Qy 368 nProValAlaThrGluArgPheGlyThrValAlaValAsnPheGlnSerSerThrAs 388
Db 418 TGGTCTCCAGACC-----TTCCAGGGCTCCAGTGTCTCGCAGCAAGACGCAAGCCCGT 365
Qy 388 pProAlaThrGlyAspValHisAlaMetGlyAlaLeuProGlyMetValTrpGlnAspAr 408
Db 364 CCCCTGCTGGAGATGTTTCATGCA-----GAAGTACATGCGGTAA 323
Qy 408 sAspValTyrLeuGlnGlyProIleTrpAlaLysIleProHisThrAspGlyHisPheHi 428
Db 322 TGATGTACACGTACAGGTTCCTCGGCTTCATGAACATCGCTCGGTGGGGGCTGCC 263
Qy 428 sProSerProLeuMetGlyLysPheGlyLeuLysAsnProProGlnIleLeuLys 448
Db 262 GGCCACCCCTTGAGTGGGGCGCTTCATCTCTGG-----CCCCAGGTATGCTC-- 214
Qy 448 sAsnThrProValProAlaAsnProProAlaGluPheSerAlaThrLysPheAlaSerPh 468
Db 213 -----GGTCTCCAGATCGGCGCTCGGAGTCTGTGCCAT 179
Qy 468 e-----IleThrGlnTyrSerThrGlyGlnValSerValGluIl 481
Db 178 TAGGAAGTCGGGACTAGGACCTGTCCAGAAATGCCCGGCCAGGGGACTGCGGCT 119
Qy 481 eGluTrpGluLeuGlnLysGluAsnSerLys 491
Db 118 GGTGGAAGAACTCCAAACCCAGTCGGGTAAG 88

RESULT 12
BO682489/c
LOCUS
DEFINITION BO682489 986 bp mRNA linear EST 15-JUL-2002
5', mRNA sequence.
AGENCY 8034509 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6213998
ACCESSION BO682489
VERSION BO682489.1 GI:21795168
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 986)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue procurement: DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLCM2380 row: j column: 15
High quality sequence stop: 612.
Location/Qualifiers
1. 986
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6213998"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_112"
/notes="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC Library."

BASE COUNT 221 a -297 c 298 g 170 t
ORIGIN

Alignment Scores:
Pred. No.: 3.47 Length: 986
Score: 106.50 Matches: 56
Percent Similarity: 36.65% Conservative: 36
Best Local Similarity: 22.31% Mismatches: 97
Query Match: 3.66% Indels: 62
DB: 13 Gaps: 8

US-09-807-802A-17 (1-534) x BQ682489 (1-986)

Qy 249 AsnGlnSerGlySerAlaGlnAsnLysAspLeuLeuPheSerArgGlySerProAlaGly 268
Db 733 CATAGAACGGAGGGGTTTCGGGGCCCACTCCCTGCATGGCCGACGCCACCCGGGCTG 674
Qy 269 MetSerValGlnProLysAsnTrpLeuProGlyProCysTyrArgGlnGlnArgValSer 288
Db 673 CTGCACATACACCGCGCTCTGCTGGCTCCAGGGACCACTCCAGCCATACAGCTTCAT 614
Qy 289 LysThrLysThrAspAsnAsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsn 308
Db 613 CCTGTGCAGTCCCTCTGCTCAAGGCTCTGTGTGATGGCCAGG----- 569
Qy 309 LeuAsnGlyArgGluSerIleLeuAsnProGlyThrAlaMetAlaSerHis-LysAspAs 328
Db 568 -----CCTGGCACAGCTTGGAGGGGCCACTGCGAGAG-- 538
Qy 328 pGluAspLysPhePheProMetSerGlyValMetIlePheGlyLysGluSerAlaGlyAl 348
Db 537 -----CTCGCGTCTCTTGGAGCACAGCTGGCGG 509
Qy 348 aSerAsnThrAlaLeuAspAsnValMetIleThrAspGluGluLeuLysAlaThrAs 368
Db 508 TGCC-----TTTCCGGAGGGTGTCTGCGAAGCTGACGCA 476
Qy 368 nProValAlaThrGluArgPheGlyThrValAlaValAsnPheGlnSerSerThrAs 388
Db 475 TGGTCTCCAGACC-----TTCCAGGGGCTCCAGTGTCTCGCAGCAAGACGCCGCCGT 422
Qy 388 pProAlaThrGlyAspValHisAlaMetGlyAlaLeuProGlyMetValTrpGlnAspAr 408
Db 421 CCCCTGCTGGAGATGTTTCATGCA-----GAAGTACATCGCGTAA 380
Qy 408 sAspValTyrLeuGlnGlyProIleTrpAlaLysIleProHisThrAspGlyHisPheHi 428
Db 379 TGATGTACACGTACAGGGTCCCGGCTCCATGAACATGCTCGGTTCGGGGGCTGCC 320
Qy 428 sProSerProLeuMetGlyLysPheGlyLeuLysAsnProProGlnIleLeuLys 448
Db 428 sProSerProLeuMetGlyLysPheGlyLeuLysAsnProProGlnIleLeuLys 448

Db 319 GGCACCCCTTGAGTGGGGCTTCATCTCTGG-----CCCCAGGTATGCTC--- 271
 Qy 448 sAenThrProValProAlaAsnProProAlaGluPheSerAlaThrLysPheAlaSerPh 468
 Db 270 -----GGTCCACGATGCGGCTCGGAGTTCTGTGCCAT 236
 Qy 468 e-----lIleThrGlnTyrSerThrGlyGlnValSerValGlu1 481
 Db 235 TAGGAAGTCCGGAGTACTGAGACTGTCCAGAAATGCCCGGCCAGGGGACTCGCGGCT 176
 Qy 481 eGluTrpGluLeuGlnLysGluAsnSerLys 491
 Db 175 GGTGGAAGACTCCACCCAGTCTGGGTAAG 145

RESULT 13
 BU800802
 LOCUS SJF2CVC08 SJF Schistosoma japonicum cDNA, mRNA sequence. EST 12-FEB-2003

BU800802
 BU800802.1 GI:28357809

EST.
 Schistosoma japonicum

Schistosoma japonicum

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;

Strigeida; Schistosomatidae; Schistosomatidae; Schistosoma.

1 (bases 1 to 606)

Sun, W., Yan, Q., Shen, D., Liu, F., Xu, X., Zhu, Z., Zhang, X., Wang, J.,

Hu, J., Xu, X., Wang, Z., Zeng, L., Rong, Y., Wu, X., Qu, J., Xu, Z.,

Huang, J., Ma, Y., Wang, S., Wang, Z., Xue, C., Feng, Z., Chen, Z. and Han

Z.

Expressed sequence tags from female adults of Schistosoma japonicum

Unpublished

Contact: Zengqiang Han

Chinese National Human Genome Center at Shanghai

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801919 (ex.45)

Fax: 86-21-50801922

Email: hanzq@chgc.sh.cn

Location/Qualifiers

1. .606

/organism="Schistosoma japonicum"

/mol_type="mRNA"

/db_xref="taxon:6182"

/sex="female"

/tissue_type="Whole body"

/dev_stage="adult"

/lab_host="rabbits"

/clone_lib="SJF"

BASE COUNT 242 a 89 c 40 g 230 t 5 others

ORIGIN

Alignment Scores:

Pred. No.: 2.37 Length: 606

Score: 105.00 Matches: 51

Percent Similarity: 36.29% Conservative: 35

Best Local Similarity: 21.52% Mismatches: 61

Query Match: 3.61% Indels: 90

DB: 13 Gaps: 14

US-09-807-802A-17 (1-534) x BU800802 (1-606)

Qy 111 AsnPhelyleuPheAenlleGlnValLysGluValThrThrAsnAspGlyValThrThr 130

Db 18 AACTATAATATCTCAATCTTCTTTA---CATACATCATCGTCATCATCAATCAACACA 74

Qy 131 IleAlaAsnAsnLeuThrSerThrValGlnValPheSerAspSerGluTyrGlnLeu--- 149

Db 75 ATAGTACCATCATATATCATCT-----TATCTCATTCACCGTTAGTTTAAAT 122

Qy 150 -----ProTyrValLeuGlySerAlaHisGlnGlyCysLeuProProPheProAlaAsp 167

Db 123 CCATTTCCATATATATTAATTTTCAAAAAAATCT-----TCAGAT 164

Qy 168 ValPheMetIleProGlnTyrGlyTyrLeuThrLeuAsnAenGlySer----- 183
 Db 165 TTTTATTATTAAACAAAAAATCAATATTATTTCTACAAATGGTTCATATCATNNNNA 224
 Qy 184 -----GlnAlaValGlyArgSerSerPheTyrCysLeuGlu----- 195
 Db 225 GAATTAATACGTACATTACAAAGTCGTAATCATTTATGTTGTTTACAAAAATTTGAATTTA 284
 Qy 196 -----TyrPhePro-----SerGln 200
 Db 285 TTTCACCATTTTCAATTTTATTTTCCATTAATTTAATTAATTTTCTACACAT 344
 Qy 201 MetLeuArgThrGlyAsnAsnPheThrPheSerTyrThrPheGlu----- 215
 Db 345 TTATATACATAATAATAATTAATTTCTACACATTTAAATGATTTTATGTGAT 404
 Qy 216 -----GluValProPheHisSerSerTyrAlaHis 225
 Db 405 TTACCAGAAATTGAAATATTGAATCGTCGACGTATACGATATCATCGT----- 452
 Qy 226 SerGlnSerLeuAspArgLeuMetAsnProleuLeuAspGlnTyrLeu---TyrTyrLeu 244
 Db 453 -----CAACAATATGTAACATTTATGTT 476
 Qy 245 AsnArgThrGlnAsnGlnSerGlySerAlaGlnAsn-----LysAspLeuLeuPhe 261
 Db 477 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 536
 Qy 262 SerArgGlySerProAlaGlyMetSerValGlnProLysAsnTrpLeuProGlyProCys 281
 Db 537 TCACAT---CCTCAAGCAACATTTCTGTAAT----- 566
 Qy 282 TyrArgGlnGlnArgValSerLysThrLysThrAspAsnAsnAsnSerAsn 298
 Db 567 TATCAACACCAACAA-----CAACAATAATAATAATAATAATAATAATAATAATAATA 605

RESULT 14

LOCUS B1253826/c

DEFINITION 602976195F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5115479 5',

ACCESSION B1253826

VERSION B1253826.1 GI:14805634

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 775)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: Incyte Genomics, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM11280 row: n column: 24

High quality sequence stop: 701.

location/Qualifiers

1. .775

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5115479"

/tissue_type="cervical carcinoma cell line"

/lab_host="DH10B"

/clone_lib="NIH_MGC_12"

source

/note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library prepared by Life
Technologies."

BASE COUNT 172 a 225 c 244 g 134 t
ORIGIN
Alignment Scores:
Pred. No.: 4.93 Length: 775
Score: 103.50 Matches: 55
Percent Similarity: 36.65% Conservative: 37
Best Local Similarity: 21.91% Mismatches: 97
Query Match: 3.56% Indels: 62
DB: 12 Gaps: 8

US-09-807-802A-17 (1-534) x B1253826 (1-775)

Qy 249 AenGlnSerGlySerAlaGlnAsnLysAspLeuLeuPheSerArgGlySerProAlaGly 268
Db 621 CATAGACCGGAGGGGTTTCGGGGCCCACTCCCTGCAATGGCCGACGCCACCCCGGGGTG 562
Qy 269 MetSerValGlnProLysAsnTrpLeuProGlyProCysTyrArgGlnGlnArgValSer 288
Db 561 CTGCCACTACAGCGGCTCACTGGGCTCCAGGGGACCACTCCAGCCATACAGCTTCAT 502
Qy 289 LysThrLysThrAspAsnAsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsn 308
Db 501 CCTGTGCCAGGTCCTCTGGTCAAGACTCTGTTGATGCCAGG----- 457
Qy 309 LeuAsnGlyArgGluSerIleIleAsnProGlyThrAlaMetAlaSerHis-LysAspAs 328
Db 456 -----CCTGGCACAGCTTGAGGGGCCACTGCGAGAG-- 426
Qy 328 pGluAspLysPhePheProMetSerGlyValMetIlePheGlyLysGluSerAlaGlyAl 348
Db 425 -----CTCGCGTCTTGGAGCACGGCTGGCGG 397
Qy 348 aSerAsnThrAlaLeuAspAsnValMetIleThrAspGluGluIleLysAlaThrAs 368
Db 396 TGCC-----TTTCCGGAGGGGTGCTGCAAGCTGACGCA 364
Qy 368 nProValAlaThrGluArgPheGlyThrValAlaValAsnPheGlnSerSerThrAs 388
Db 363 TGCTCTCAGACC-----TTCCAGGGCTCAGTGTCCGAGCAAGACGACCCCGT 310
Qy 388 pProAlaThrGlyAspValHisAlaMetGlyAlaLeuProGlyMetValTrpGlnAspAr 408
Db 309 CCCCCTGGCTGGAGATGTTTCATGCA-----GAAGTACATCCGCTAAA 268
Qy 408 sAspValTyrLeuGlnGlyProIleTrpAlaLysIleProHisThrAspGlyHisPheH 428
Db 267 TGATGTACACGTACAGGGTCCCGGCTTCATGAACATGCTCGTGGTGGGGGGTCTGCC 208
Qy 428 sProSerProLeuMetGlyGlyPheGlyLeuLysAsnProProGlnIleLeuIleLy 448
Db 207 GGCCACCCCTTGAGTGGCGGCTTCATCTCTGG-----CCCAAGGTATGCCCTC-- 159
Qy 448 sAsnThrProValProAlaAsnProProAlaGluPheSerAlaThrLysPheAlaSerPh 468
Db 158 -----GGTCTCCAGATCGCGGCTCGGAGTCTCTGGCAT 124
Qy 468 e-----IleThrGlnTyrSerThrGlyGlnValSerValGluI 481
Db 123 TAGGAAGTCGCGGACTAGGACCTGTCCAGAAATGCCCGGCCAGGGGAGTCTCCGGCT 64
Qy 481 eGluTrpGluLeuGlnLysGluAsnSerLys 491
Db 63 GGTCAAGAACTCCAAACCCAGTCGGGTAAG 33

RESULT 15
BF340815 865 bp mRNA linear EST 22-NOV-2000
LOCUS
DEFINITION 602037709P1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4185483

5', mRNA sequence.

ACCESSION BF340815
VERSION BF340815.1 GI:11287279
KEYWORDS EST.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE

1 (bases 1 to 865)
NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished

CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-rc@mail.nih.gov

Tissue Procurement: David N. Louis, M.D.
CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM9505 row: a column: 04

High quality sequence stop: 585.

Location/Qualifiers

1. 865

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4185483"

/tissue_type="glioblastoma with EGFR amplification"

/lab_host="DH10B (TI phage-resistant)"

/clone_lib="NCI_CGAP_Brn64"

/notes="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.57 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 249 a 260 c 222 g 134 t

ORIGIN

Alignment Scores:

Pred. No.: 5.82 Length: 865

Score: 103.50 Matches: 47

Percent Similarity: 34.58% Conservative: 27

Best Local Similarity: 21.96% Mismatches: 93

Query Match: 3.56% Indels: 47

DB: 10 Gaps: 6

US-09-807-802A-17 (1-534) x BF340815 (1-865)

Qy 296 AsnSerAsnPheThrTrpThrGlyAlaSerLysTyrAsnLeuAsnGlyArgGluSerIle 315
Db 73 AACGAGCAGTTTAACCTGGGTGTCCCGGCTGGCAACCTCAGCGAAGCGAAGACCGATAC 132
Qy 316 IleAsnProGlyThrAlaMetAlaSerHisLysAspAspGluAspLysPhePhePromet 335
Db 133 TATCTCGGGTCCACCGTGGTGTCCACACTTCTGACTCGGACGTTCTTCGGGTGTC 192
Qy 336 SerGlyValMetIlePheGlyLysGluSerAlaGlyAlaSerAsnThrAlaLeuAspAsn 355
Db 193 ACTGAGTGTGCTGAGCTCTTTGACTCT-----GATCCC 228
Qy 356 ValMetIleThrAspGluGluIleLysAlaThrAsnProValAlaThrGluArgPhe 375
Db 229 ATCATGTGACGCTCCCTGTAGAAGTCTCCAGGAGAACCCCTAAATTTATGGAGACCGTG 288
Qy 376 GlyThrValAlaValAsnPheGlnSerSerSerThrAspProAlaThrGlyAspValHis 395
Db 289 GCGGAGAAAGCGTGCAGGAATACCCCAAAAGACCCGG----- 327
Qy 396 AlaMetGlyAlaLeuProGlyMetValTrpGlnAspArgAspValTyrLeuGlnGlyPro 415
Db 328 -----AGGAGTGAG 336

